UNIVERSITY OF CALIFORNIA
Santa Barbara

Spatiotemporal Granulation

A Dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Media Arts and Technology

by

Muhammad Hafiz Wan Rosli

Committee in Charge:

Professor Curtis Roads, Chair

Professor Clarence Barlow

Professor JoAnn Kuchera-Morin

Dr. Andres Cabrera

Dr. Matthew Wright

June 2017

The Dissertation of
Muhammad Hafiz Wan Rosli is approved:

Professor Clarence Barlow

Professor JoAnn Kuchera-Morin

Dr. Andres Cabrera

Dr. Matthew Wright

Professor Curtis Roads, Committee Chairperson

June 2017

Spatiotemporal Granulation

by

Muhammad Hafiz Wan Rosli

Dedicated to my parents,

Wan Rosli Wan Daud & Zarita Zainuddin

Acknowledgements

Family

Thank you very much for your boundless support and encouragement. I could not have done it without each and every one of you: Zarita Zainuddin, Wan Rosli Wan Daud, Muhammad Huzaimi Wan Rosli, Saarah Huurieyah Wan Rosli, Muhammad Huzaifah Wan Rosli, Haseenah Huurieyah Wan Rosli, and Muhammad Hanzalah Wan Rosli.

Media Arts and Technology (UC Santa Barbara)

Committee

I would like to extend my deepest gratitude to my supervisor, Professor Curtis Roads. Your friendship and guidance has been instrumental in shaping my personal and academic life. My sincere appreciation goes out to my committee members, Professor Clarence Barlow, Professor JoAnn Kuchera-Morin, Dr. Andres Cabrera, and Dr. Matthew Wright for their patience and confidence in my research. It has been a great pleasure to work with you.

Educators and Staff

For the years of knowledge that has been bestowed upon me, thank you Professor Marko Peljhan, Professor Marcos Novak, Professor George Legrady, Professor Theodore Kim, Professor Scott Marcus, and Don Howell. I would also like to thank the hardworking staff who have provided support and assistance in handling bureaucratic matters: Laura Cheung, Lesley Fredrickson, Kris Listoe, Lisa Thwing, and Larry Zins.

Colleagues

This document has greatly benefited from suggestions and criticism provided by MAT students and community members, including Anis Haron, Timothy Wood, Owen Campbell, Akshay Cadambi, Joseph Tilbian, Karl Yerkes, Danny Bazo, Keehong Youn, F. Myles Sciotto, Alexis Crawshaw, Joshua Dickinson, Xarene Eskandar, Fernando Rincon Estrada, Charlie Roberts, Graham Wakefield, Haru Ji, Pablo Colapinto, and Juan Escalante.

Universiti Sains Malaysia

I am deeply indebted to Universiti Sains Malaysia for the Academic Staff Training Scheme fellowship. Thank you to Associate Professor Omar Bidin who has been very kind, and understanding throughout the whole process. My appreciation goes out to mentors and colleagues in Malaysia: Hasnul Saidon, Norfarizah Bakhir, Che Mat Jusoh, Jufri Yusoff, Nur Zaidi Azraai, Aswari Mohd Salleh, Zuriawati Ahmad Zahari, Md Sany Md Hanif, and Nurhaslina Mohd Kamsin.

Hochschule für Gestaltung (HfG)

Thank you to the Spatial Audio Group of HfG for co-organizing the *Space-Media-Sound* exchange program. The usage of equipment, and criticism has positively affected this research. My appreciation is extended towards Dr. Paul Modler for his support during the time spent in Karlsruhe. My gratitude is also extended towards Lorenz Schwarz and Marco Kempf for their kindness, and friendship. I would also like to thank all the people I met in Germany for their generosity and hospitality.

Zentrum für Kunst und Medientechnologie (ZKM)

Thank you to the Institute for Music and Acoustics of ZKM, especially Ludger Brümmer for the use of equipment, in particular during the developmental phase of *Angkasa*. My gratitude is also extended towards Götz Dipper and Marie-Kristin Meier who kindly assisted, and streamlined the necessary processes required in acclimatizing to the institution.

External Review

I am grateful for the valuable external feedback from John Chowning, Jean-Claude Risset, Diemo Schwarz, Barry Truax, Markus Noisternig, Natasha Barrett, Ioannis Zannos, Chandrasekhar Ramakrishnan, Jonatas Manzolli, Pedro Rebelo, and Marcelo Queiroz. Thank you for sharing your thoughts and suggestions.

Funding: Malaysian Ministry of Education

Thank you for the generous support given throughout my graduate studies. It would have been much more difficult without.

Funding: Baden Wüttemberg Foundation (Stipendium)

Thank you for providing support during the *Space-Media-Sound* exchange program.

Funding: Mosher Foundation (Fellow)

My gratitude goes out to the Mosher Foundation for their monetary support.

Funding: University of California Institute for Research in the Arts (Grant)

My appreciation is extended towards UCIRA, and Professor Kim Yasuda for partly sponsoring this research.

Muhammad Hafiz Wan Rosli- Curriculum Vitae

Date of Birth 13 September 1983
Nationality Malaysian
Telephone +1 (805) 252-2006
Email hafiz@mat.ucsb.edu

Education

2011-Present Ph.D. Media Arts and Technology - University of California, Santa Barbara
Expected date of graduation: Spring 2017
Dissertation: *Spatiotemporal Granulation*
Committee: Curtis Roads (Chair), Clarence Barlow, JoAnn Kuchera-Morin, Andres Cabrera, Matthew Wright

2008-2010 M.F.A Computer Arts - School of Visual Arts, New York
Recognition: Paula Rhodes Memorial Award

2004-2007 B.F.A New Media Arts - Universiti Sains Malaysia, Penang
Recognition: Dean's list (all semesters)

Employment History

Academic

2017 Media Arts and Technology, UC Santa Barbara, CA, United States
Teaching Assistant- MAT276N: Special Topics in Electronic Music- Modular Synthesis

Advanced topics in computer music composition, synthesis, and digital signal processing.

Technologies: Modular Synthesizer

2016 Department of Art, UC Santa Barbara, CA, United States
Instructor- Art122PC: Advanced Digital Topics- Physical Computing

The course focuses on the development of New Media projects through the exploration of open-source computer software and hardware development tools. This session, the class works on a collaborative project for Isla Vista LightWorks.

Technologies: Arduino, Autodesk 123D, Fritzing, Pure Data

2016 Department of Art, UC Santa Barbara, CA, United States
Teaching Assistant- Art7D: Art, Science and Technology

Foundations of digital and technological arts in all forms, including the history, theory and practice of kinetic, interactive, interdisciplinary, network and systems-oriented art

Technologies: HTML/ Web development

2015 Department of Art, UC Santa Barbara, CA, United States
Instructor- Art122PC: Advanced Digital Topics- Physical Computing

The course focuses on the development of New Media projects through the exploration of open-source computer software and hardware development tools.

Technologies: Arduino, Processing, Pure Data

2015 Dos Pueblos High School, CA, United States
Mentor: Dos Pueblos Engineering Academy

Design new models of education based on science, technology, engineering, art, and mathematics (STEAM). Mentor high school students for the development of interactive projects for the Maker Faire.

Technologies: Arduino, Computer Numerical Controlled Fabrication

2012 Media Arts and Technology, UC Santa Barbara, CA, United States
Teaching Assistant- MAT200A: Intersections of Art and Technology

The course is designed for arts-engineering interdisciplinary collaborative work with a focus on problem-solving within the context of media arts.

Technologies: Project-dependent

Non-Academic

2015 The Center for Research in Electronic Art Technology, CA, United States
CREATE Technical Coordinator

- Laboratory, studio, and performance space management for all CREATE and MAT Audio/Visual facilities
- Weekly maintenance as well as planning, design and installation of upgrades
- Setup and takedown of concerts and lectures
- Day-to-day support of instructional facilities including security and student keycodes

2013 Art, Design and Architecture Museum, UC Santa Barbara, CA, United States
Freelance: Exhibition Designer

- Assist artists with setting-up of artworks
- Design custom framing and installation for artworks

2012 Media Arts and Technology, UC Santa Barbara, CA, United States
System Admin Assistant

- Server maintenance
- Technical support for students

Publications

Journal Muhammad Hafiz Wan Rosli, Andres Cabrera: *Gestalt Principles in Multimodal Data Representation.* IEEE Computer Graphics and Applications 03/2015; 35(2): 80-87.DOI:10.1109/MCG.2015.29

Proceeding Muhammad Hafiz Wan Rosli, Curtis Roads: *Spatiotemporal Granulation.* International Computer Music Conference, Utrecht, Netherlands; 09/2016

Muhammad Hafiz Wan Rosli, Andres Cabrera: *Angkasa: A Software Tool for Spatiotemporal Granulation.* International Symposium on Computer Music Multidisciplinary Research, University of São Paulo, Brazil; 07/2016

Muhammad Hafiz Wan Rosli, Andres Cabrera, Matthew Wright, Curtis Roads: *Granular Model of Multidimensional Spatial Sonification.* Sound and Music Computing, Maynooth University, Ireland; 07/2015

Muhammad Hafiz Wan Rosli, Karl Yerkes, Timothy Wood, Hannah Wolfe, Charlie Roberts, Anis Haron, Fernando Rincon Estrada, Matthew Wright: *Ensemble Feedback Instruments.* New Intefaces for Musical Expression, Baton Rouge, LA, US; 05/2015

Muhammad Hafiz Wan Rosli, Andres Cabrera: *Application of Gestalt Principles to Multimodal Data Representation.* IEEE VIS Arts Program, Paris, France; 11/2014

Selected Exhibitions

2016 *Architecture of Life.* Berkeley Art Museum, California, US
2014 *REKA: International Conference on Creative Media, Arts and Technology.* Universiti Sains Malaysia, Penang, Malaysia
2014 *IEEE VIS Arts Program.* Paris, France

2014 *Sound and Music Computing + International Computer Music Conference.* Athens, Greece
2014 *Media Arts and Technology End of Year Show.* UC Santa Barbara, California, US
2013 *New Interfaces for Musical Expression.* Daejeon/ Seoul, South Korea
2012 *Soundwalk, East Village Arts District.* Los Angeles, California, US
2012 *Bits and Pieces.* UC Santa Barbara, California, US
2010 *Rooms Art Uncovered, issue 02 (2010): 032-033*
2010 *Maker Faire, New York Hall of Science.* New York, US
2010 *MFA Computer Arts Thesis Exhibition.* Visual Arts Gallery, New York, US
2010 *User Generated.* School of Visual Arts, New York, US
2010 *LUKIS.* USM, Penang, Malaysia
2007 *Tabir Maya.* Tuanku Fauziah Museum and Gallery, Penang, Malaysia
2005 *Upstart: The Nokia Creative AV Awards.* Kuala Lumpur, Malaysia

Selected Performances

2017 *Angkasa.* MAT End of Year Show, UC Santa Barbara, California, US
2017 *Modular Synthesis.* MAT End of Year Show, UC Santa Barbara, California, US
2016 *CREATE Ensemble.* CCRMA March Mattness, Stanford University, California, US
2016 *CREATE Ensemble.* CREATE Allosphere Concert, UC Santa Barbara, California, US
2015 *CREATE Ensemble.* HfG Eroeffnung des Wintersemesters 2015/2016, Karlsruhe, Germany
2015 *CREATE Ensemble.* inSonic2015, Karlsruhe, Germany
2015 *CREATE Ensemble.* New Interfaces for Musical Expression, Baton Rouge, Louisiana, US
2015 *Gibber Live Coding.* Algorave, Santa Barbara, California, US
2015 *Siamang.* MAT End of Year Show, UC Santa Barbara, California, US
2014 *CREATE Ensemble.* Autumn Waveforms, UC Santa Barbara, California, US
2014 *CREATE Ensemble: World premiere of Feedback.* UC Santa Barbara, California, US
2014 *CREATE Ensemble: Des Gestes Touchants.* California, US
2014 *CREATE Ensemble: Medium, Ensemble Live Coding.* UC Santa Barbara, California, US
2013 *CREATE Ensemble: Ambiguous Suggestions.* UC Santa Barbara, California, US
2013 *CREATE Ensemble: Twykr: Study of Transtemporal Ensemble Instrument.* UC Santa Barbara, California, US

Selected Workshops

2012 *The Art of Gathering Environmental Data: Hybrid Sensor Network Workshop.* Helsinki, Finland
2012 *Ars Bioarctica.* Kilpisjarvi, Finland (Funded by Projekt Atol, C-TASC, Finnish Bioart Society)

2012	*Unmanned Resilience: Sensor Network Workshop.* Gora Plateau, Slovenia
2012	*Sensing Resilience: SiNuNi sensor network workshop.* Linz, Austria

Professional Services

2016	Student Representative: MAT, UC Santa Barbara, California, US
2015	inSonic Conference Workshop Chair: ZKM/ HfG, Karlsruhe, Germany
2015	inSonic Conference Local Organizing Committee: ZKM/ HfG, Karlsruhe, Germany

Honors and Awards

2016	Block Grant: UC Santa Barbara
2016	Grant: University of California Institute for Research in the Arts
2015	Scholarship: Baden Wuerttemberg Stipendium
2015	Exchange Program: ZKM/ HfG Space-Media-Sound
2014	Fellowship: Mosher Foundation
2011	Block Grant: UC Santa Barbara
2011	Scholarship: Ministry of Education, Malaysia Post-Graduate Studies (Ph.D.)
2010	Award: The Paula Rhodes Memorial Award
2008	Scholarship: Ministry of Education, Malaysia Post-Graduate Studies (Masters)

Affiliations

Institution	Media Arts and Technology, UC Santa Barbara
Researcher	Center for Research in Electronic Art Technology, UC Santa Barbara
Researcher	Allosphere Research Group, UC Santa Barbara
Fellow	School of Arts, Universiti Sains Malaysia
Mentor	Dos Pueblos Engineering Academy
Sponsor	Ministry of Education, Malaysia
Sponsor	Mosher Foundation
Member	International Computer Music Association
Member	Audio Engineering Society
Member	Arctic Perspective Initiative
Member	CREATE Ensemble
Member	Indian Classical Music Ensemble, UC Santa Barbara
Member	Gamelan Kyai Selamet Ensemble, UC Santa Barbara

Languages

Malaysian	Arterial language
English	Fluent (Speaking and Writing)
Programming	*Arduino*, *C++*, *HTML*, *LaTeX*, *MAX/MSP*, *Processing*, *Pure Data*, *Python*

Other Interests

Travel
Scuba (Rescue Diver Certification)
Sailing (Keelboat Sailing Certification)

Referees

Name	Prof. Curtis Roads
Affiliation	MAT/ Music, UC Santa Barbara
Position	Professor (PhD Supervisor)
Contact	clang@mat.ucsb.edu

Name	Prof. Clarence Barlow
Affiliation	MAT/ Music, UC Santa Barbara
Position	Professor (PhD Committee)
Contact	barlow@music.ucsb.edu

Name	Dr. Matthew Wright
Affiliation	CCRMA, Stanford University
Position	Technical Director (PhD Committee)
Contact	matt@ccrma.stanford.edu

Abstract

Spatiotemporal Granulation

Muhammad Hafiz Wan Rosli

This document introduces a novel theory and technique called Spatiotemporal Granulation. Through the use of spatially encoded signals, the algorithm segments spatial and temporal information, producing grains that are localized in both space and time. Well-known transformations that are derived from classical granulation, as well as new manipulations that arise are discussed, and outlined.

As a means to reassemble the grains into a new configuration, we explore how granulation parameters acquire a different context, and introduce new methods for control. We present findings and limitations of this new technique, and outline some potential creative and analytical uses. The viability of Spatiotemporal Granulation is demonstrated through a software implementation titled *Angkasa*.

Contents

List of Figures

"...it must now be remembered that there is an infinite number of spaces in motion with respect to one another."

(Einstein 1952)

Chapter 1

Introduction

The process of segmenting a sound signal into small grains (less than 100 ms), and reassembling them into a new time order is known as granulation [59]. Various techniques can be used to articulate the spatial characteristics, allowing one to choreograph the position and movement of individual grains. This spatial information, however, is generally synthesized, i.e., artificially generated, unlike temporal information which can be extracted from the sound sample itself, and then used to drive resynthesis parameters.

Ambisonics [33] is a technology that represents full-sphere spatial sound (periphonic) information through the use of Spherical Harmonics. This research aims to use spatial information extracted from the Spherical Harmonics as a means to granulate space.

By extracting this spatial information, the described method creates novel possibilities for manipulating sound. It allows the decoupling of temporal and spatial information of a grain, making it possible to independently assign a specific position in time and space for analysis and synthesis.

1.1 Background

Classical granulation[1] segments a one-dimensional signal into grains lasting less than 100 ms, and triggers them algorithmically. As opposed to the means of artificially generating spatial information, we are interested in extracting grains from different positions in space (akin to extracting grains in time for classical granulation), and using the embedded information for synthesis.[2]

By granulating (segmenting) the spatial domain, in addition to the temporal domain of a captured signal, we would be able to extract grains that are localized in space and time. The ability to do so would allow us to reassemble the grains in a new spatial and temporal configuration, as well as introduce a range of possibilities for transformation.

[1] *Classical Granulation* is a term used to differentiate Granulation [59], from *Spatiotemporal Granulation*

[2] The known techniques used to spatialize grains are described in Section 1.6

1.2 Problem Definition

An encoded spatial sound, such as those captured using Ambisonics, contain spatial information of a sound scene. The encoded sound scene can then be decoded over a variety of setups, including stereo, quadrophonic, pluriphonic (multiple loudspeakers), or rendered into binaural format using a *Head Related Transfer Function* [47].

Prior to the decoding process, various transformations such as Rotate, Tilt, Tumble, Focus, Warp, or effects such as global reverberations allow us to manipulate the encoded spatial scene– but only in its entirety [42]– the spatial representation must remain *intact*. In other words:

A sound in a specific spatial and temporal position can only exist in its original position in space and time, relative to other sounds.

This stands in contrast to classical granulation (of stored sound files), where grains from any time-point can be extracted, and altered in temporal arrangement, i.e., scrambled in time (Microsound, 2001, p 192).

1.3 Research Questions

> *The representation of a signal purely in terms of its time series (e.g., the sequence of samples of a discrete signal) or purely in terms of its frequency content (Fourier expansion) were both extreme cases of a wide range of signal expansions. They show high locality (precision) in one domain, but no locality at all in the other. For the analysis and manipulation of sound, good locality in both time and frequency is desirable* (Gabor 1946, quoted in Roads 2001)

The Gabor Transform proved that any sound could decompose into a combination of (temporally) localized Fourier Transforms. As an extension to Gabor's premise:

1. Is it possible to decompose spatial sound into a combination of acoustical quanta (grains) in time, frequency, and space?

    1.1. Is it possible to use the inherent encoded spatial information from spatial sounds for the decomposition?

    1.2. Are these grains unique in time, frequency, and space?

2. Is it possible to select and trigger the grains in a different order?

    2.1. Is it possible to freeze the time domain, and trigger the localized grains in space?

3. Is it possible to reassemble the grains into new spatial and temporal configurations?

3.1. What are the known granular transformations that could be adopted?

3.2. What novel effects are introduced?

1.4 Solution

The solution to the problem presented above is to introduce an approach to decompose (discretize/ segment) an encoded spatial sound into a combination of elementary acoustical *quanta* [27, 59], bounded in time, frequency, and space. Spatiotemporal Granulation aims to address this problem using the inherent spherical harmonics encoded in Ambisonic recordings. Segmentation of spatial, temporal, and frequency domains lead to a number of novel transformations and effects, including:

1. Micro Manipulation of sound trajectory: Change the trajectory of a moving sound object

2. Relativity effects: Space-time cavity– allowing different parts of the spatial scene to progress at different time speeds

3. Micromontage [59] in space and time

4. Manipulate, transform, and apply effects to only certain micro sections of the spatial scene

5. Spatial filtering: Select a portion of space to be synthesized– select/ remove contents based on spatial position, as opposed to frequency domain filtering

6. Spatial Scanning: "Scan" the captured space, independent from the time domain, i.e., spatially exploring a moment frozen in time

7. Spatial source separation: Source separation based on spatial location

8. Source width control: Change apparent size of sound object

9. Remap one sound's spatial configuration to another sound

10. Feature analysis of grains in space

11. New descriptor for describing space, based on analyzed features

12. Rearrange grains based on features, similar to concatenative synthesis [66], but for both time and space

13. Spatial cross-synthesis: Substitute selected grains with other grains from a different spatial position, temporal position, or from different sound sources

1.5 Purpose

The purpose of this research is to investigate if it is indeed possible to segment the spatial and temporal domains of a spatial recording, using the embedded

spatial information. The outcome is a theory and technique that proves the above, and an implementation that demonstrates the theory– It allows a user to perform spatiotemporal granulation for analytical, and creative uses, described in Section 5.

The scope of this research is limited to sonic materials that contain encoded spatial information. The spatialization component will be carried out using the *AlloSphere*, as well as rendered into *Binaural format*. The justifications for these limitations are discussed in Section 1.7.1.

1.6 Related Work

The segmentation of spatial audio, and the extraction of grains from different spatial and temporal positions, is a research area that had not previously been explored. However, there has been a number of techniques used to synthesize sound particles in space (spatialization). These techniques are discussed in Section 2.4.1.

1.7 Methodology

The fundamental basis for the theory of Spatiotemporal Granulation is the segmentation of spatial sounds into grains that are localized in time, frequency, and space. The theory will be verified through experiments carried out using a software implementation, described in Section 5. Due to the limitations discussed in Section 1.7.2, only spatial audio materials captured using *Ambisonics* will be considered in the experiments. These experiments include:

1. **Analysis:**

    1.1. Decompose spatial sound, and prove that the spatiotemporal grains are unique in time, frequency, and space

2. **Transformation:**

    2.1. Extract the spatiotemporal grains, and selectively trigger them in any arbitrary order

3. **Synthesis:**

    3.1. Reassemble the grains into new spatial and temporal patterns (manipulate the spatial sound's encoded configuration)

        3.1.1. Resynthesis: Reconstruction from extracted grains

        3.1.2. Complex multi-channel reassembly

1.7.1 Limitations

The spatialization aspect of this research is inherently limited by a (multichannel) system's ability to render sounds in space. The AlloSphere [7] is a unique facility that is capable of rending spatial sounds using 54 loudspeakers in a spherical (pluriphonic) configuration.

It is arguably the world's leading spatialization instrument, and therefore will be the primary source for determining the success of criteria (3) in Section 1.7. Other multichannel systems, such as University of Birmingham's BEAST, Simon Fraser University's AudioBox, and *Zentrum für Kunst und Medientechnologie*'s Klangdom will not be used in the current context of this research. These systems will be explored in future developments.

Additionally, the AlloSphere also provides 360° real-time stereographic visualization using a cluster of servers driving 26 high-resolution projectors. This would allow each spatiotemporal grain to be acoustically, and visually localized in its corresponding location [63]. The multimodality in perceiving visual and auditory stimuli (Section 3.1.5) would assist us to further investigate the Research Questions, allowing us to better contrast the original encoded sound field, and the transformed sound field.

Furthermore, the re-encoded spatial sound will be binaurally rendered [47] as a means to simulate the spatialization effect. Doing so would allow us to further

assess criteria (3) in Section 1.7 as an additional measurement, independent of the AlloSphere.

1.7.2 Delimitations

There are several microphone technologies that allow the capturing of spatial audio information, such as *X-Y/ Blumlein Pair*, *Decca Tree*, and *Optimum Cardioid Triangle*. However, these technologies do not capture the complete full-sphere information of spatial sound.

Ambisonics recording is a technique that captures periphonic spatial information via microphone arrays, such as the "SoundField Microphone" [33]. It is important to note that this technique equally treats sounds from any direction, as opposed to other techniques that assumes the frontal information to be the main source, and other directional information as ambient sources. Thus, the scope of this research is limited to spatial sounds encoded using Ambisonics. [3]

1.8 Data Collection

During the fourth quarter of 2015, the author attended a research residency/ exchange program called "Space-Media-Sound" at the *Staatliche Hochschule für*

[3] As new spatial sound recording technologies emerge, this research will be extended to include them, but will focus on Ambisonics for now.

(a) Captured space

(b) Close up of the ST350

Figure 1.1: Example of microphone setup

Gestaltung (Karlsruhe University of Arts and Design), and *Zentrum für Kunst und Medientechnologie* (Center for Art and Media). The facilities that were available at both institutions were used for data collection and software implementation. We would like to extend our gratitude towards the institutions mentioned above for their kind assistance in this part of the research.

The data used in this research were primarily captured using the *SoundField ST350 surround microphone*, and processed on a *2015 Mac Pro (OSX 10.10.5)* via an *RME Fireface UFX audio interface*. The recordings were captured from various venues around Europe (Figure 1.1), and the United States, containing materials such as speech, natural environments, and sounds of insects.

Additionally, B-Format files were also downloaded from `www.ambisonic.net`, `www.ambisonia.com`, and `www.ambisonic.info`. Plans to capture using Higher-order microphone arrays are currently in progress.

Chapter 2

Microsound

> *Below the level of the musical note lies the realm of Microsound, of sound particles lasting less than one-tenth of a second. Recent technological advances allow us to probe and manipulate these pinpoints of sound, dissolving the traditional building blocks of music – notes and their intervals – into a more fluid and supple medium. The sensations of point, pulse (series of points), line (tone), and surface (texture) emerge as particle density increases. Sounds coalesce, evaporate, and mutate into other sounds.* (Roads 2003)

2.1 History

The fundamental idea that sounds can be decomposed into elementary building blocks in time can be traced back to the ancient philosophies of atomism, dating back to ancient India [30], and ancient Greece. The Western world was introduced to this theory in the fifth century BC, when Leucippus and Democritus speculated that any substance could be successively divided into smaller pieces, until it is no

longer divisible– the basic element– the atom. The atomistic philosophy was comprehensive: both matter and energy (such as sound) were composed of tiny particles [59].

Atomism was revived at the dawn of early modern science, which gradually forced a paradigm shift away from Aristotelianism, into a more experimental perspective. In 1616, Isaac Beekman proposed a "corpuscular" theory of sound, and speculated that any vibrating object cuts the surrounding air into spherical particles that project in all directions, only to be perceived as sound as these particles arrive at the eardrum [10, 59].

Formulation of atomic theory continued throughout the years, but the particle theory of sound was opposed due to the belief of sound as a wave phenomenon [31, 59]. In 1907, Albert Einstein predicted that ultrasonic vibration could occur on the quantum level of atomic structure (verified in 1913), leading to the concept of *acoustical quanta*, or *phonons* [24, 59].

The scientific development of a quantum or granular approach to sound was first proposed by the British physicist Dennis Gabor in three papers that combined theoretical insights into quantum physics (1946, 1947, 1952), with practical experiments, including the Electro-optical and Electromechanical Sound Granulator. In contrast to the timeless description (infinite duration signals) of Fourier Analysis, Gabor's solution involved the combination of frequency, and time, i.e.,

any sound can be decomposed into a combination of elementary grains, bounded by frequency, and time [27, 28, 29]. Today, one refers to the analyses that are limited to a short time frame as a windowed analysis, for example, localized Fourier transforms, such as the *Short Term Fourier Transform*. Dennis Gabor's research influenced other researchers of the time, including Werner Meyer-Eppler, Abraham Moles, and Norbert Wiener [60].

Iannis Xenakis was a prominent composer that was well aware of developments in the scientific world, and this knowledge influenced his musical theories. He hypothesized that every sound could be understood as an "assemblage of a large number of elementary sounds adequately disposed in time" [59]. His compositions *Metastasis*, and *Concret PH* were clear examples of statistical and granular processes. These works do not involve strict mathematical operations, but were rather derived intuitively by the composer. On the other hand, *Analogique B* consists of grains that were generated from cutting a tape that contains sound recorded using analog tone generators. The rules of the composition involved scattering grains on to grids called *screens*, which represented elementary sonic quanta bounded by frequency, amplitude, and time.

Karlheinz Stockhausen (assisted by Gottfried Michael Koenig) in 1960 composed *Kontakte* using only filtered impulses, generated by Analog Impulse Generators. He formulated a theory between infrasonic frequencies, and the audible

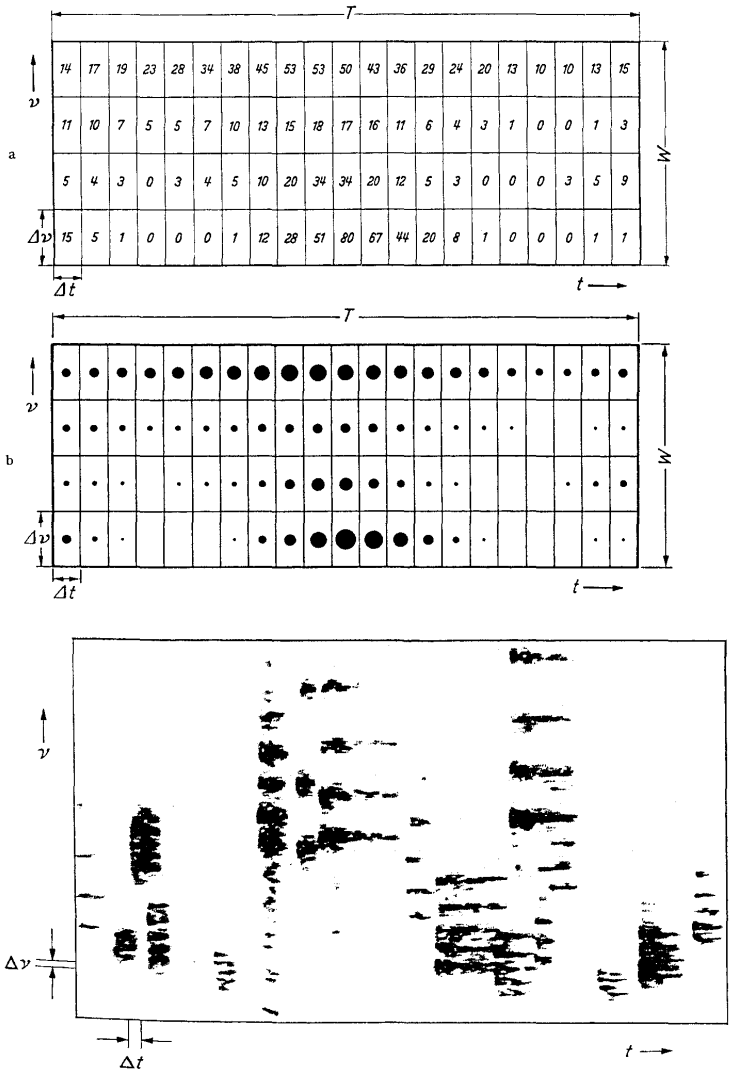


Figure 2.1: The Gabor matrix. The top image indicates the energy levels numerically. The middle image indicates the energy levels graphically. The lower image shows how the cells of the Gabor matrix (bounded by Δv, where v is frequency, and Δt,, where t is time) can be mapped into a sonogram. From Roads (2001).

frequencies of impulses. A detailed analysis of relationships between musical time scales is presented in his text *". . . . . How time passes . . . . . "*. Stockhausen's *"The unity of musical time"* summarizes his integrated approach to composition, and describes the procedure used to blur the boundaries between rhythm and pitch in his composition *Kontakte*.

Curtis Roads is arguably the leading researcher in Granular Synthesis, having written numerous texts on the subject, including *Microsound* [59]. His music has been produced by MIT Media Laboratory, Wergo, OR, Mode, and Asphodel, and performed in various world renowned festivals. In 1974, he developed the first computer based implementation of Granular Synthesis, written in Extended Algol and Music V, followed by a study entitled *Prototype*. As a Research Associate at the Experimental Music Studio (Massachusetts Institute of Technology), he developed 2 forms of Granular Synthesis in the C Programming Language: a sinusoidal granular synthesis engine, and a sampled sound granulator using Music 11. After the release of personal computers, he went on to program new implementations of granular synthesis (*Synthulate*) and granulation of sampled soundfiles (*Granulate*) using the Apple Macintosh II. These implementations became part of the combined source code to form an interactive application called *Cloud Generator*, developed with John Alexander while working at Les Ateliers UPIC in 1995.

Today, we have numerous implementations of Granular Synthesis on various platforms. For a comprehensive overview of the systems and implementations, please refer to Microsound, p. 112.

2.2 Theory

Granular Synthesis is a sound synthesis technique that operates on elementary building blocks called grains. A grain is a brief microacoustic event, approaching the minimum perceivable event time for duration, frequency, and amplitude discrimination (Whitfield 1978; Meyer-Eppler 1959; Winckel 1967; Roads 2001). It is typically within the range of 1 to 100 ms in duration, and contain a waveform shaped by an envelope (Figure 2.2). The grain can either contain synthetic materials such as a sine tone, or extracted from a captured signal (Section 2.3). It belongs to a family of sound particles, where each particle is defined by its envelope type, waveform, and characteristics. Table 2.1 summarizes the variety of sound particles.

Name	Envelope Type	Waveform	Characteristics
Grains	Gaussian or arbitrary	Arbitrary, including sampled	Can be scattered irregularly or metrically; each grain is potentially unique; Gaussian grains are compatible with analysis systems like the Gabor transform and the phase vocoder
Glissons	Gaussian or arbitrary	Arbitrary	Variable frequency trajectory; synthesizes glissandi or noise textures
Pulsars	(1)Rectangular around pulsaret, then null; (2)Gaussian (3)Expodec (4)Arbitrary	Arbitrary	Independent control of fundamental and formant spectrum; can also be applied in the infrasonic frequencies as a rhythm generator; synchronous distribution of pulsars
Trainlets	Expodec or arbitrary	Impulse	Used to synthesize tones; offers independent control of fundamental and formant spectrum;
Wavelets	Gaussian or other, subject to mathematical constraints	Sinusoidal	Particle duration varies with frequency; starts from an analysis of an existing source sound
Grainlets	Gaussian or arbitrary	Sine	Interdependent synthesis parameters
Micro-arcs	Arbitrary	Arbitrary, including sampled	Flexible graphic design; subject to graphical transformations
FOF grains	Attack, sustain, release	Sine	Envelope controls formant spectrum
Vosim grains	Linear attack, exponential decay	Sine^2 pulses	Used for pitched tones; flexible control of spectrum
Window-function pulses	Rectangular around pulse then nil	Blackman-Harris pulse	Synthesizes formants with purely harmonic content
Transient drawing	Arbitrary	Hand-drawn	Generally sharp and percussive; each transwave is unique
Particle cloning	Arbitrary	Arbitrary, including sampled	Repeats a particle so that it becomes a continuous tone

Table 2.1: Sound particles. From Roads (2001).

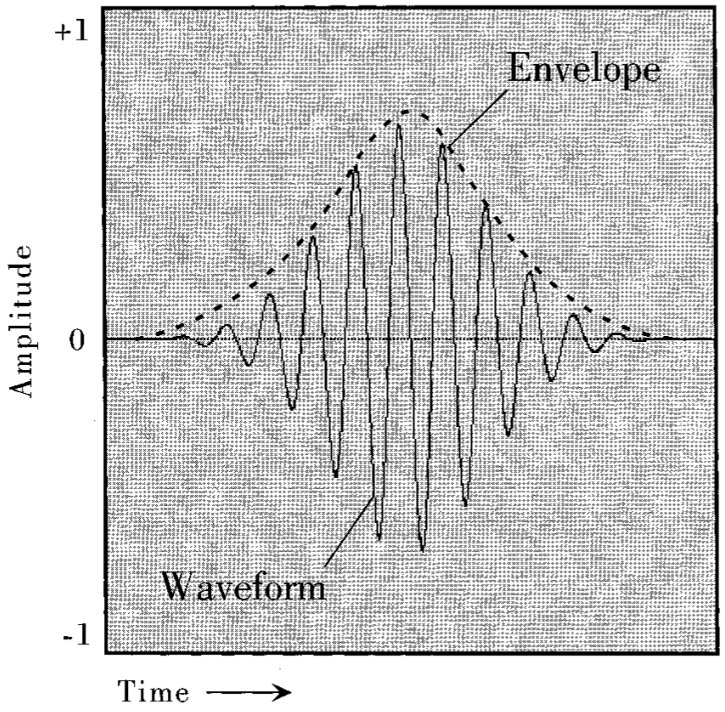


Figure 2.2: Portrait of a grain in the time domain. The duration of the grain is typically between 1 and 100 ms. From Roads (2001).

The main forms of Granular Synthesis can be divided into six types, based on how the grains are organized. They are as follows [59]:

1. Matrices and screens on the time-frequency plane

2. Pitch-synchronous overlapping streams

3. Synchronous and quasi-synchronous streams

4. Asynchronous clouds

5. Physical or abstract models

6. Granulation of sampled sound

A grain generator is a simple synthesis instrument which contains a wavetable oscillator, shaped by an envelope. Although the instrument may be simple in principle, the sound output's complexity is derived from each grain's varying parameters. For a comprehensive list of microsound synthesis techniques, please refer to Microsound, pp. 92-116, 119 - 178, 253-299.

Individual grains that are less than 2 ms resemble acoustical clicks, but a combination of the grains create a cloud texture. Minor variations in parameter of each grain causes strong changes in the overall spectrum of the sound output. The following are a selection of parameters that shape the sonic material:

2.2.1 Grain Envelope

Each grain is shaped by an amplitude envelope.[1] The grain envelope creates an amplitude modulation (AM) effect, generating sidebands around the carrier frequency of the grain at intervals of the envelope period. If the grain duration is D, the center frequency of the AM is 1/D. The grain envelope can take any shape, and can vary based on other parameters, such as in the case of the wavelet transform, and grainlet synthesis. Example of the different types of windows are:

[1] In Dennis Gabor's original conception, the envelope used was a Gaussian window

1. *Bell-shaped Gaussian curve* smoothest envelope on a mathematical standpoint (Figure 2.3(a))

2. *Quasi-gaussian* retains the smooth attack and decay but has a longer sustain portion in the envelope– increased perceived amplitude (Figure 2.3 (b))

3. *Simple line segment* practical reasons– save memory space and computation time (Figure 2.3 (c))

4. *Band-limited pulse or sinc function* imposes a strong modulation effect (Figure 2.3 (e))

5. *Expodec* percussive attack articulates rhythmic structure (Figure 2.3 (f))

6. *Rexpodec* long attack envelope with a sudden decay. Granulated concrete sounds appear to be "reversed" when played with rexpodec grains, even though they are not (Figure 2.3 (g))

2.2.2 Grain Duration

Grain duration is an important parameter in determining the sound output of a Granular Synthesis engine. Grains with a short duration (less than 5 ms) does not give the impression of pitch, while grains longer than 25 ms gives a clearer pitch sensation.

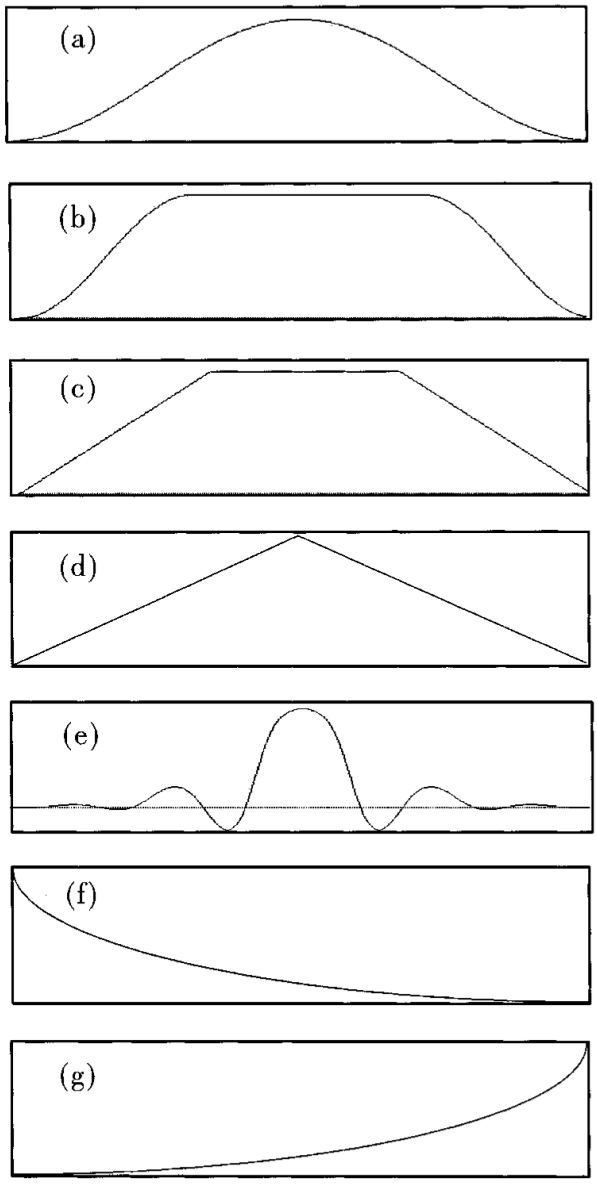


Figure 2.3: Grain envelopes. (a) Gaussian. (b) Quasi-Gaussian. (c) Three-stage line segment. (d) Triangular. (e) Sinc function. (f) Expodec. (g) Rexpodec. From Roads (2001).

Estimation of the optimum grain duration varies from 10 ms to 60 ms. The laws of micro-acoustics tell us that the shorter the duration of a signal, the greater its bandwidth [59]. In order to completely represent one period of a given frequency, the grain duration must be at least equal to the frequency period. If we take this rule into account, the minimum length of a grain would be 50 ms (for a 20 Hz signal).

However, it is indeed possible for the duration of a grain to be less than 50 ms for a 20 Hz signal, but this short grain creates modulation products. For example, grains shorter than 5 ms creates particulated clouds where a sense of center-pitch is still present, but is diffused by noise. There are four classes of grain durations within a cloud [59]:

1. *Constant duration* the duration of every grain in the cloud is constant
2. *Time-varying duration* the grain duration varies through time
3. *Random duration* the duration of a grain is random
4. *Parameter dependent duration* the duration of a grain is tied to certain parameters

2.2.3 Grain Waveform

A grain can contain any type of waveform, be it simple fixed waveforms such as the sine, saw, square and sinc; or varying waveforms on a grain-by-grain basis. Using different waveforms in a single cloud leads to cloud *color type* [58]. These color types can be categorized as *monochrome* (contains a single waveform), *polychrome* (contains two or more waveforms), and *transchrome* (evolution of waveform over the duration of the cloud). Alterantively, the grain waveform can also be extracted from a sampled sound.

2.2.4 Frequency Band Effects

This parameter limits the fundamental frequency of grain waveforms. The grain generator scatters grains within the upper and lower limit of this parameter. It creates either a complex texture from grains with random frequency distribution, or a harmonic texture from grains constrained to a certain set of frequencies within a scale.

2.2.5 Density and Fill Factor

The density parameter corresponds to the number of grains per second. Both grain density and grain duration define the resulting texture. The fill factor of a cloud is defined as the product of a cloud's density and its grain duration. For

clouds with varying grain density, we derive the average density and average fill factor based on the mean between two extremes.

In the case of Asynchronous Granular Synthesis, the triggering of grains are not sequential. Some grains may overlap, creating silence at other points in the cloud. A good rule of thumb to create a *solid cloud* is to set the density per second to at least 2/ grain duration (in seconds). Increasing the grain density results in differences in the texture, notably [59]:

1. Narrow bands and high densities generate pitched streams with formant spectra

2. Medium bands (e.g., intervals of several semitones) and high densities generate turgid colored noise

3. Wide bands (e.g., an octave or more) and high densities generate massive clouds of sound

2.2.6 Granular Spatial Effects

The scattering of grains in one spatial location creates a flat spatial perspective. However, as grains are spread across the sound field by assigning individual positions, we gain a three-dimensional spatial representation. The perception of spatial resolution is governed by the physical properties of the signal, and by

the human auditory system's ability to distinguish stimuli in different locations– Localization blur [16]. Localization blur is the phenomena where a sonic point source creates an auditory image that spreads out in space (further discussed in Section 3.1.2). Various techniques for spatializing the grains are described in Section 2.4.

2.3 Granulation

Granulation is the process of segmenting a sound signal into small (less than 100 ms) grains. These grains may be further modified through various means of transformations, and reassembled into a new time order and microrhythm.

Granulation can either be performed purely in the time-domain, or in both time and frequency domains. Techniques such as Fourier, Wavelet and Gabor transforms analyzes the spectrum of each grain after they are temporally segmented. Because granulation can accept any type of signal as an input, the possibilities of the sonic result are boundless. Thus, the output is greatly dependent on the input signal, and parameters for the granulation engine. These parameters are controllable via numerical script, physical controllers, or deterministic/ stochastic algorithms. Examples of granulation parameters include [59]:

1. Selection order– from input stream: sequential (left to right), quasi-sequential, random (unordered)
2. Pitch transposition of the grains
3. Amplitude of the grains
4. Spatial position of the grains
5. Spatial trajectory of the grains (effective only on large grains)
6. Grain duration
7. Grain density– number of grains per second
8. Grain envelope shape
9. Temporal pattern–synchronous or asynchronous
10. Signal processing effects applied on a grain-by-grain basis–filters, reverberators, etc.

Asynchronous playback from a sound file allows one to extract individual grains in any arbitrary order: sequential, reversed, random, statistical evolution, or based on certain features. For example, we can extract grains with a certain spectral distribution (e.g., of a specific instrument), in any time point of the sound file. We can then play these grains in the order that we extracted them, instead of

its original sequence. Furthermore, we can extract grains from a different instrument, and interweave the grains together. Alternatively, we can create a spectral evolution of clouds from the first instrument to the second instrument.

Selective granulation is the process of separating different components of a sound signal, and granulating only certain analyzed components. Once the grains are extracted, we can perform an analysis of each grain, and selectively synthesize grains based on certain qualities, or features. Alternatively, we can apply effects and transformations only to certain grains based on the analysis, for example apply a filter with a unique center frequency and bandwidth for each grain.

Real-time granulation can either be performed via an incoming sound source, or by using a stored sound file. This technique allows us to extract a grain, and play the same grain repeatedly a number of times before going on to the next grain, causing the sound output to be stretched with a factor equivalent to the number of times the grain is played back. Conversely, we can skip grains to speed up the sound. Increasing or reducing the playback sampling rate while doing the aforementioned process allows us to shift the pitch of the output signal without changing its duration.

Overlaying multiple copies of a grain with different phase delays increases its perceived volume and creates a kind of chorus effect. By varying the shape and size of the granulation window, we are able to distort the input sound in a

controllable way. One of the drawbacks of real-time granulation is that the read pointer is unable to look ahead in time. This makes it impossible to perform time-compression, or time-scrambling.

Convolution of microsounds allow us to "marry" two signals [54], creating a new signal that is sonically related to both input signals. Examples of the most striking effects include attack smoothing, multiple echoes, room simulation, time smearing, and reverberation. Although the promise of convolution on the micro level is very attractive, caution has to be taken as it can easily destroy the identity of both sources.

2.4 Granular Spatialization

Spatial sound and its history in electronic music will be thoroughly discussed in Chapter 3. In this section, we present spatialization specifically in the context of microsound, and the current practices.

The scattering of grains in one spatial location creates a flat spatial perspective. For example, when a dense cloud of grains are played back on a single channel (monaural), the entire cloud is collapsed to a single point in space. In contrast, if each grain is assigned an individual position, and spread across the sound field, we gain a three-dimensional spatial representation. Although this effect is better perceived with an array of loudspeakers placed in certain config-

urations (Section 3.3.2), an addition of one loudspeaker to the monaural output makes a big difference to the spatial perspective.

Our ability to localize sounds is governed by the physical properties of the signal, and by the human auditory system's ability to distinguish stimuli in different locations. The Duplex Theory by Lord Rayleigh [56] states that the directionality of sound sources is dependent on two mechanisms– The Interaural Time Difference, and The Interaural Intensity Difference. The former is our ability to measure the location and distance of a sound source using the time difference between our two ears, while the latter is our ability to measure the location and distance of a sound source using intensity differences. It is important to note that these two measurements are unique to each individual, which forms the basis of HRTF[2].

Our ability to locate sound sources is also dependent on localization blur [16], discussed in Section 3.1.2. This phenomenon exists because a point source creates an auditory image that spreads out in space, varying based on the content of the signal, and the location of the sound source in relation to the listener.

In spatializing the sound sources, the precedence effect (also known as *law of the first wavefront*) has to be taken into account. Analogous to forward masking [59], where two closely spaced successive tones are perceived as one, the precedence effect states that listeners perceive a single fused auditory image when a sound is

[2]Humans are also able to learn new patterns, and use them as a new measurement tool [36]

followed by another sound in a short period of time. More importantly, the sound's perceived spatial location is dominated by the location of the first arriving sound.

The phenomena described above will be discussed further in Chapter 3.

2.4.1 Current Practice

There has been a number of research related to the process of positioning sound particles in space– spatialization. The following are a selection of current practices.

C. Roads, Microsound. MIT Press, 2001, pp. 221-234

Roads outlines the techniques used for spatialization of microsound into two main approaches [59]:

1. **Scattering of sound particles in different spatial locations and depths**. Grains can be placed in one spatial position, or spread out across multiple positions. While it is indeed possible to manually assign individual position for grains over multiple loudspeakers via a Digital Audio Workstation (sound editor/ mixing program), assigning a large number of grains over a large number of loudspeakers becomes rather complicated, if not impossible. As a solution, automatic scattering algorithms provide a means to

position the grains using higher level control. For example, the *Cloud Generator* program offers different options for scattering grains in a stereo sound field, including panoramic motion, and random distribution. The complexity of this process is dependent on the number of grains, and number of channels for playback. As more and more pluriphonic spaces are developed, new spatialization algorithms will become available to accommodate the different layouts and configurations (Section 3.3.2).

2. **Using sound particles as spatializers for other sounds via granulation, convolution, and intermodulation**. The basis of this technique is to use other sources as spatializers, instead of manual or algorithmic processes. For example, one can extract the (multi-channel) amplitude envelope of generated synthetic particles, and impose it on to another source signal. Convolution of a sound source with a cloud of grains is another means of spatialization (Figure 2.4). Analogous to convolution of sounds with a room Impulse Response (IR), the grains can be thought of as individual IRs of a weird environment [59]. For low density clouds, the result is a statistical distribution of echoes (asynchronous), or metrical rhythms resembling tape echo (synchronous). As the grain density increases, the echoes fuse into a quasi-reverberation effect. If each grain contains a single sample, the re-

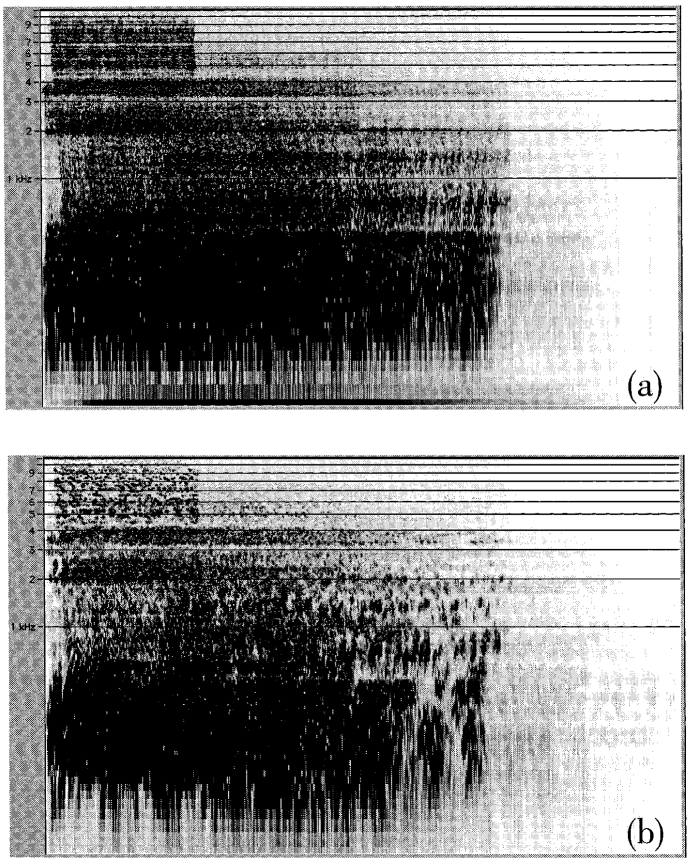


Figure 2.4: Spatialization via convolution with sound particles. These sonograms are the results of convolutions of a vocal utterance with two dense clouds of particles. The sonograms used a 2048-point FFT with a Kaiser-Bessel window. Frequency is plotted logarithmically from 40 Hz to 11.025 kHz. (a) The particle envelope is expodec (sharp attack, exponential decay). (b) The particle envelope has a Gaussian attack and decay. Notice the turgid undulations caused by time-smearing due to the smooth attack. From Roads (2001).

sult would be echoes of the input signal. On the other hand, if each grain contains multiple samples, a time-smearing effect is attained.

N. Barrett, "Spatio-musical composition strategies", Org. Sound, vol. 7, no. 3, pp. 313-323, Dec. 2002

Barrett describes four notions of space, and discusses how a composer can work with these ideas. They are as follows:

1. **llusion of a space or a spatial location of an object.** The author discusses creating a spatial illusion by maintaining real 'spatial laws', such as the effect of sound transmission, the properties of reverberant field, object image size and multiple object relationship, and doppler shifts. A combination of these aspects are what determines the spatial illusion, whether the sound-object is recognizable or abstract.

2. **Allusion to a space or a spatial location of an object.** Spatial allusion happens when there is no direct connection to the interrelated acoustic laws mentioned above. Under these situations, the 'listening imagination' is important. The author suggests that these conditions are perhaps where the 'music' begins to emerge– The transformation from illusion to allusion.

3. **Simulation of the three-dimensional (3D) sound field.** Barrett's work focuses on the use of ambisonics techniques– from individually encoding

grains with spatial information via Higher-order Ambisonics (HOA) to a hybrid approach of combining first order recorded sources and HOA synthesized materials. She states that the main differences in composing in stereo versus pluriphonic are:

(a) The technical aspect of mixing and sound transformations– The technical aspect of sound reproduction includes the selection of closely matched loudspeakers, the placement, and the decoding process. Sound transformations need to be carried out *before* the material is spatially positioned, as transformations will often destroy the encoded space (except for granulation, where the transformation is incorporated in the encoding process).

(b) Issues of clarification in spatial information– The sound location, and gestural spatial definition is enhanced through the use of pluriphonic systems. This results in clearer (real and surreal) 'landscapes', and clearer presentation of object and spatial relationships.

(c) Additional compositional considerations– Pluriphonic systems increase the number of simultaneously perceptually identifiable sounds. There is a more intimate relationship with the listener through sound proximity, and allows for convincing reproduction of real-world sound en-

vironments. As a consequence, one is able to create consonance and conflict in the 3D sound field.

4. **Spatial possibilities contingent upon temporal development.** In the context of contemporary electronic music, new motifs in pitch, rhtyhm and timbre are introduced through the unfolding of a piece. In contrast, a listener's perception of spatial patterns exist through connecting with real-world experience. When a composer transforms the spatial information, the temporal information may depart from any real world experience, which would require listeners to train their memory through the act of listening, similar to temporal motifs– The composer can challenge the listener with unique spatial information.

B. Truax, "Composition and diffusion: space in sound in space", Organised Sound, vol. 3, pp. 141-146, Aug. 1998

Truax [73] discusses the manipulation of spectral and temporal shape of a sound object in the practice of timbral composition. Truax links the process of timbral design to spatialization, focusing on the decorrelation of sound sources (and its components) as a primary factor for determining both the perceived magnitude and spatial form (spatial distribution of grains in space). Granular time-stretching, he states, is perhaps the single most effective approach, as it

contributes to the spectral richness, duration, and unsynchronized temporal components.

The overlaying of several unsynchronized streams, coupled with delays decorrelates each granular stream. These independent streams are then sent to individual loudspeakers, without the use of spatialization algorithms. The essence of his argument is to create significant relationships between shaping the *perceived volume of the sound (internal space)* and the *distribution of sound in space (external space)* by combining the two processes using a single algorithm.

D. Kim-Boyle, "Spectral and granular spatialization with boids", in ICMC 2006, International Computer Music Conference, 2006

Kim-Boyle introduces a novel way of spatializing grains in space using flocking algorithms, specifically Craig Reynold's *boids* [57] algorithm. Reynold's original factors for determining the behavior of the flock are separation, alignment, and cohesion. The three parameters refer to the preferable distance one bird would maintain from another, the tendency of a bird to fly towards the average heading of its local neighbors, and the tendency of a bird to steer towards the average position of its local neighbors. These parameters are used in Eric Singer's *boids* object for MAX/MSP.

The same parameters, as well as those that are introduced by Eric (such as intertia of the boids, willingness to change speed and direction) are later used by Kim-Boyle as a means to position individual grains. Additionally, he experimented with defining the trajectory of motion using simple circular trajectories, and modification of trajectories with performance data. Transformation of the boid's movement includes the ability to add an offset to a coordinate, squish the coordinates into a particular region in space, apply a separate rotational force to the coordinates, and map regions in space into which the boids will not fly.

In addition to using the algorithm to spatialize grains, Kim-Boyle has also mapped the movement of boids to spatial location of individual bins in a Short-Term Fourier Transform, similar to Torchia and Lippe's work on frequency-domain based spatial distribution [72], and the author's work on granular model of multidimensional spatial sonification [64]. Similar work based on flocking algorithms has been carried out by Blackwell & Young (Swarm Granulator) [15], and Wilson (Spatial Swam Granulation) [77].

E. Deleflie and G. Schiemer, "Spatial-grains: Imbuing granular particles with spatial-domain information", in ACMC09, The Australasian Computer Music Conference, July 2009

The techniques outlined above aim to position grains in a particular location in space– spatialization. On the other hand, Etienne Deleflie and Greg Schiemer proposed a technique to encode grains with spatial information extracted from an Ambisonics file. As the spatial encoding is defined by the relationships between the combined components, the authors proposed that the spatial encoding is retained if the micro-control for each grain is maintained.

This technique is highly dependent on the original encoding, and the explored spatial opportunities is limited to the source material. Granulation is performed on the time domain (temporal segmentation), and each grain is accessed by specifying a *time position* in the source ambisonics file. In other words, time is used as an index to the library of spatial information contained within the source. The authors described 2 implementations written in Pure Data [52], and SuperCollider [43], which allowed the synthesis of non-point sound sources, such as a line, planar, and volumetric renderings.

2.4.2 Spatialization on Multiple Timescales

The macroscale spatial perspective dominated most of the early works in electronic music. Global reverberation is a characteristic attribute of the time, which can be seen in Oskar Sala's *Elektronische Impressionen* (1978). Stockhausen's *Kontakte* in 1960 showed a more diverse spatial impression, contrasting the foreground and background relationships through selective reverberation.

The next step to this can be seen as positioning individual sound objects in a three dimensional space. Articulation of the space can be explored through the unfolding of phrases, while contrapuntal relations can be derived via relationships between stationary and moving objects. Going down to the building blocks of sound, spatial organization can also be carried out on the micro scale. For example, as discussed in Section 2.4.1, Truax spatializes 8 granular streams to 8 distinct loudspeakers, while Roads positions individual grains in space via *granular spatial scattering*. Using this technique, grains are spatially distributed in granular form, giving each grain a unique position in space, while retaining all other aspects of the sound, such as pitch, duration and timbre. The Creatovox synthesizer, developed by De Campo and Roads in 2003, not only scattered each grain in a unique spatial position, but also applies a unique per-grain reverberation over an octophonic sound system in real time.

Spectrum analysis techniques decompose a sound into time-frequency (TF) representations. An algorithm searches for specific features in the TF representation, such as transient events, loud components, short components, and spatialize these events based on a predetermined set of rules. The *Dictionary Based Pursuit* (DBP) decompose a sound into grains that are localized in time and frequency [70]. Different properties of the grain are extracted, and spatialization is performed based on these properties. For example, grains with a short duration would be positioned in a different location compared to grains with a long duration. Spectral Modeling Synthesis (SMS), The Tracking Phase Vocoder (TPV), and basic Fast Fourier Transform (FFT) could also be used to generate a similar effect as DBP. In the case of SMS, one could analyze the features of the TF representation, and spatialize them independently. For spectrum analysis via FFT, each band can be positioned in a separate virtual position [64]. Analysis-based spatialization methods might sound promising, but they face challenges in determining appropriate sounds, and issues of interactive control. Similar to most experimental methods, these techniques requires testing and tuning to get the most compelling results [60].

Chapter 3

Spatial Sound

Our perception of the world is dominated by the visual senses. The other senses (auditory, tacticle, etc) are much less developed in comparison, which results in a more visual-centric description of our environment. Blauert [16] makes a distinction between the physical phenomena that are characteristic of sound[1] events, i.e., the physical aspect of sound; and the perceptual aspect of sound (perceived component), as in the term "auditory event". This distinction is made to emphasize the fact that not every auditory event is connected with a sound event, i.e., one does not necessarily lead to another in a causal relationship. The

[1]The German Standard DIN 1320 (1959) defines *sound* as "mechanical vibrations and waves of an elastic medium, particularly in the frequency range of human hearing (16 Hz to 20 kHz). However, a more recent definition of sound, published by the American National Standards Institute (ANSI) defines it as "(a) Oscillation in pressure, stress, particle displacement, particle velocity, etc., propagated in a medium with internal forces (e.g., elastic or viscous), or the superposition of such propagated oscillation. (b) Auditory sensation evoked by the oscillation described in (a)."

connection between the two events are by no means direct, and the auditory event is not necessarily perceived as sharing the same location as the sound event.

3.1 Spatial Hearing

> *Sound events and auditory events are distinct in terms of time, space, and other attributes* (Lungwitz 1933b)

Hearing is inherently spatial in nature. This fundamental premise is derived from the fact that sound and auditory events occur *only* at particular times, at particular places, and with particular attributes. Every auditory event has its own unique spatial characteristic, based on the relationship between the locations of auditory events and other parameters related to the physiology of the brain [16]. These spatial parameters can convey meanings and sensations in ways that other senses are unable to. For example, auditory events may occur at positions where nothing is visible– behind the observer, in the dark, blocked by a physical object, etc. In most cases, the auditory event and the sound event shares the same location in space. However, this is not necessarily true, as we will discuss later on in the text. Specifically, we will see how spatialization allows us to synthesize *phantom sources* and place them at different locations, and not necessarily where the vibrating body (loudspeaker) resides.

Blauert carries out systematic investigations on the physical and psychological phenomena that effects spatial hearing, dividing the study into cases with a single sound source, and multiple sources. Although the single sound source instance represents an elementary case for analysis from the standpoint of physics (linearity of the equations for sound fields) [16], the situation becomes more complicated when analyzed from a psychophysical view (non-linearity of nervous system).

3.1.1 The Auditory System

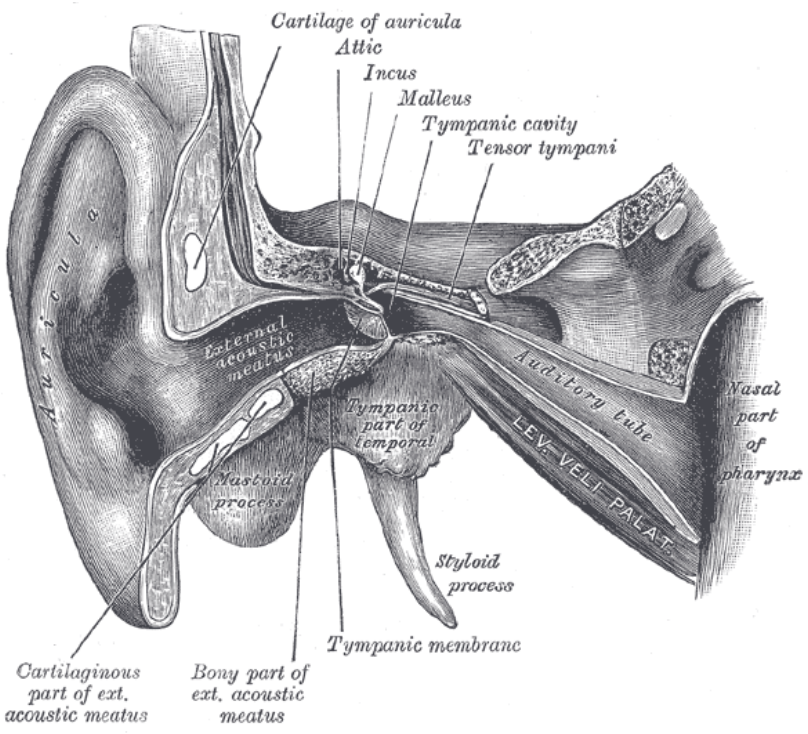


Figure 3.1: The Auditory System. From Gray (1918), Plate 907.

The auditory system (Figure 3.1) is the sensory system for the sense of hearing. Human beings have two ears located at approximately the same height on the left and right sides of the head. Each ear can be separated into 3 parts, namely:

1. **External ear (*Auris Externa*)** In the context of spatial sound, this part of the ear plays the most important role in the perception of auditory events. It consists of the pinna (*Auricula*), and the external ear canal (*Meatus Acusticus Externus*).

   The pinna is composed of a cartilage structure covered in skin, and connected to the surrounding parts by ligaments and muscles. It's form and shape is distinctly unique and varies from one individual to another. The pinna lies at the side of the head, and surrounds the entrance to the ear canal. The most basic function of the pinna is as a focusing cone, where it gathers sound energy and focuses it into the external ear canal.

   Acoustically the pinna functions as a linear filter whose transfer function depends on the direction and distance of the sound source [16]. The process of altering the input signal based on the direction and distance of the sound source makes the pinna an important, if not the most important part of spatial hearing. It codifies the spatial information of the sound field into temporal and spectral attributes. This effect is dependent on various

physical phenomena such as reflection, shadowing, dispersion, diffraction, interference, and resonance [56].

The external ear canal is a slightly curved tube which connects the hollow part of the pinna to the eardrum. It conducts the vibrations to the tympanic cavity, and amplifies frequencies in the range of 3 kHz to 12kHz. A simplified analytical model of the external ear can be stated as follows [16]: "*The pinna, along with the ear canal, forms a system of acoustical resonators. The degree to which individual resonances of this system are excited depends on the direction and distance of the sound source*".

2. **Middle ear (*Auris Media*)** The middle ear consists of the eardrum (*Membrana Tympani*), the tympanic cavity (*Cavum Tympani*), and the ossicles within the cavity, known as the hammer (*Malleus*), anvil (*Incus*), and stirrup (*Stapes*).

   The eardrum is a 0.1mm thick, slightly elliptical diaphragm, and lies at an angle of approximately 40° - 50° to the axis of the ear canal. It is the receiver of sound[2], and what divides the external ear, and the middle ear. Variations of pressure in the ear canal excites the eardrum via the ossicles mentioned above. These ossicles converts the lower pressure vibrations at the eardrum

[2]Additionally, sounds can also be received in the external ear canal via excitation of the lining of the canal– transmitted through the temporal bone to the inner ear, i.e., bone conduction

into higher pressured vibrations at another smaller membrane called the *Oval Window*. This high pressure is necessary because the inner ear beyond the oval window contains liquid rather than air. Within the middle ear, the sound information is still represented as a waveform (continuous pressure variations).

3. **Inner ear (*Auris Interna*)** The inner ear includes the Cochlea (the auditory portion of the inner ear), and the vestibular organs (which contains the receptors for the sense of balance).

   The Cochlea is a spiralled, hollow, conical chamber of bone, in which waves propagate from the base (near the middle ear and the oval window) to the apex (the top or center of the spiral) [16]. The Organ of Corti (within the Cochlea) is the sensory organ for hearing, and is distributed along the partition separating fluid chambers in the coiled tapered tube of the Cochlea. Hair cells[3] in the Organ of Corti converts the sound pressure patterns into electrochemical impulses which are passed on to the brain via the auditory nerve.

[3]Specifically the organized structure called *Stereocilia*

3.1.2 Localization and Localization Blur

Localization is defined as the rule by which the location of an auditory event (e.g., its direction or distance) is related to attributes of a sound event. In other words, it is the listener's ability to identify the location or origin of a detected sound in direction and distance.

The position of the sound source, the type of signal it radiates, and previous sound events [16, 59] can effect localization. Under certain circumstances, a single sound source can produce simultaneous auditory events, causing confusion in one's ability to localize the sound source. The ability to localize a sound source also varies from one person to another, based on physical factors [16], as will be discussed later on. The main question of this rule pertains to where an auditory event appears, in relation to a specific given position of a sound source.

Localization Blur is a property of localization. It is the smallest change (Just Noticeable Difference) in specific attributes of a sound event that renders a change in the perceived auditory event [16]. For example, the minimum difference in direction of a sound source that allows a human to perceive a change in the sound's direction. By definition, this concept establishes the fact that a human's ability to perceive spatial characteristics is less refined compared to the space that the sound source exists in. Physical measuring techniques such as microphones, allows us to capture the spatial information in a more detailed resolution, compared to a

human's auditory system. A sound event that exists in a single position in space (point source), for example, creates a spatially spread auditory event.

Based on research done by [69, 45, 16], the region with the most precise spatial hearing is around the forward direction of a listener. That is to say, the frontal region is the most sensitive to spatial difference, i.e., highest spatial resolution. Small changes in location of the sound source around the frontal direction results in clear directional changes in the auditory event. The lower limit for the localization blur is about 1°. Table 3.1 shows a survey of measurements for localization blur on the median plane (horizontal displacement) for narrow-band signals such as sinusoids and Gaussian tone bursts, as well as speech, broadband noise, and impulses.

The localization blur increases with displacement from the forward direction toward the left or right of the listener. At right angles (in relation to the sound source), the localization blur increases by three to ten times its value compared to the forward direction [16, 69, 71]. In contrast, the localization blur for backwards direction is approximately twice the value of the forward direction. Figure 3.2 shows the result of an experiment described by Preibisch-Effenberger (1966) and by Haustein and Schirmer (1970) using a large number of untrained users. 100 ms white noise pulses of approximately 70 phon were used for the experiment.

Reference	**Type of signal**	**Localization blur (approximate)**
Klemm (1920)	Impulses (clicks)	0.75° - 2°
King and Laird (1930)	Impulse (click) train	1.6°
Stevens and Newman (1936)	Sinusoids	4.4°
Schmidt et al. (1953)	Sinusoids	$> 1°$
Sandel et al. (1955)	Sinusoids	1.1° - 4.0°
Mills (1958)	Sinusoids	1.0° - 3.1°
Stiller (1960)	Narrow-band noise, $\cos^2$ tone bursts	1.4° - 2.8°
Boerger (1965a)	Gaussian tone bursts	0.8° - 3.3°
Gardner (1968a)	Speech	0.9°
Perrott (1969)	Tone bursts with differing onset and decay times and frequencies	1.8° - 11.8°
Blauert (1970b)	Speech	1.5°
Haustein and Schirmer (1970)	Broadband noise	3.2°

Table 3.1: A survey of measurements of localization blur $\Delta(\varphi = 0)_{\min}$ i.e., for horizontal displacement of the sound source away from the forward direction. Because different measuring techniques were used, reference is made to the original works, where the techniques are described in detail. From Blauert (1997).

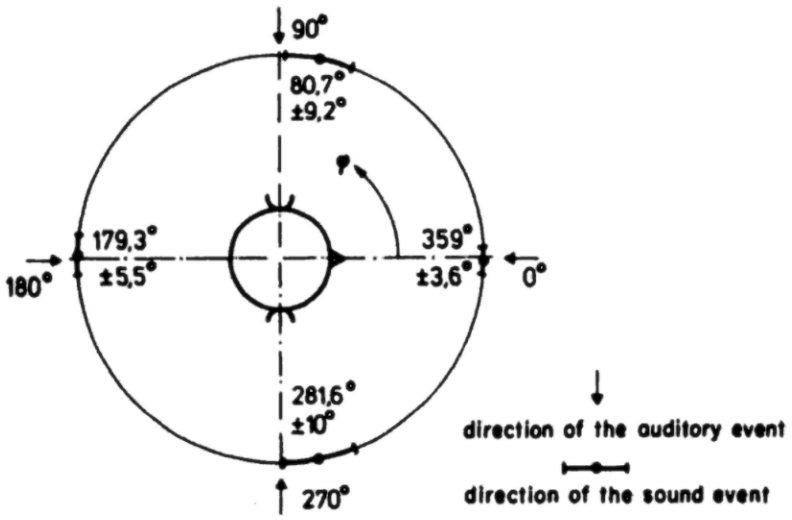


Figure 3.2: Localization blur $\Delta\varphi_{\min}$ and localization in the horizontal plane (after Preibisch-Effenberger 1966a and Haustein and Schirmer 1970; 600-900 subjects, white-noise pulses of 100 ms duration, approximately 70 phon, head immobilized). From Blauert (1997)

These results vary based on the spectral content, and duration of the signals. For a more detailed analysis, please refer to [45, 16]. It has been suggested that a single sound source with multiple narrow-band components might be perceived as several simultaneous or successive auditory events in different directions. For example, Von Hornbostel (1926) reported that a bird song appears to be changing positions, even though the bird itself is stationary. Another confusion in the process of localizing sound sources is where the auditory event is not perceived as coming from the direction of the sound source, but in a direction that is axially

symmetric with respect to the axis of the ears (line passing through both ears) [55, 69, 26]. Blauert [16] explains that this ambiguity is (in most cases) resolved if the listener is allowed to freely move their head.

Localization on the median plane is different than the horizontal plane, whereby the signals that arrive at both ears are identical, causing the sound to not be measured using the interaural signal differences. Instead, in these specific cases, localization is based on monaural cues (Section 3.1.3). The localization blur on the median plane is far greater than that of the horizontal plane, starting at 4° for white noise, and becoming greater (worse precision) for other signals such as speech [16]. This result is greatly dependent on the familiarity of the subject with the sound signal.

Additionally, localization may change as a function of time, based on the time-dependent characteristic of spatial hearing: it's persistence (or inertia). The position of the auditory event can only change with limited rapidity, and exhibits a time lag with respect to a change in position of the sound source. Persistence must always be considered when using sound sources that rapidly change position.

3.1.3 Ear Input Signals

The sound signals at the eardrums are the most important input signals for spatial hearing, and **even slight modifications to the signals can lead to noticeable differences in spatial perception**.

Duplex theory of localization

The Duplex Theory of Localization, proposed by Lord Rayleigh in 1907 states that humans (and animals) use the sound difference in time and level between the two ears to localize sound sources. The theory is based on the fact that the two ears exist on different sides of the head, and facing somewhat opposing directions. Thus, sound events arrive at different times and different amplitudes at the two ears except for sounds that occur on the median plane. Additionally, the head and pinna introduces distortions to the input signal, and functions as a filter (shadowing effect) for the difference in amplitudes (group delay).

The dissimilarities between the two ear input signals gives the auditory system enough information to determine the auditory event's location. These dissimilarities can be categorized as follows:

1. **Interaural Time Difference (ITD)** Dissimilarities between the two input signals related to time when the signals occur, or when specific parts of the sound occur. The auditory system evaluates interaural time differences

from phase delays at low frequencies, and group delay at high frequencies. The effect is greatly dependent on the content of the sound (spectrum). The auditory system can interpret time shifts in both the carrier and the envelope. Carrier time shifts primarily have an effect only below 1.6 kHz, while envelope time shifts have more of an effect for signals that contain high frequencies.

2. **Interaural Level Difference (ILD)** Dissimilarities between the two input signals related to their average sound pressure level. This is caused by head shadow (or acoustic shadow) where the sound reaching the other ear is obstructed by the head, and may have to travel around the head. This creates an attenuation of the signal and filtering effects, which gives a cue for the auditory system to localize a sound source. The difference in pressure level is effective throughout the audible frequency range. It is generally believed that the relative importance of interaural time and level differences depends on the type of sound signal. ILD have their greatest importance when the signal is composed of sounds above 1.6 kHz and the level is low.

Monaural Cues

If our understanding of spatial information is only based on the duplex theory, then sound sources in the median plane (from in front of the listener to the back)

would be perceived as sounds without differences in ITD and ILD. However, it is naturally understood that these sounds in the median plane can indeed be localized. In these specialized cases, the monaural attributes of the ear input signal provides the most significant cues in perceiving the distance and elevation angle of the auditory event. These distortions are dependent on the sound input direction and is unique to each individual. The distortions are due to the following physical effects: shadowing, reflection, and diffraction by the head and torso; reflection, dispersion, and diffraction by the pinna, and resonances in the system composed of each pinna, ear, canal, and eardrum.

Head Related Transfer Function

The Head Related Transfer Function (HRTF) is a response that characterizes how an ear receives a sound from any point in space. As described above, there are several cues that inform the auditory system about the location of sound sources. These differences in time and frequency are encapsulated within the HRTF representation. Although the definition of timbre is somewhat ambiguous[4] [18], it is clear to us (based on the above) that changes in timbre for both monaural and interaural cues affect the perception of space.

[4]The American National Standards Institute defines *timbre* as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar"

Modifications of the sound source as it reaches the eardrum may be captured using an impulse response, called the Head Related Impulse Response (HRIR). Convolution of a source signal with the HRIRs would simulate the sound a listener would hear in the original captured location. The HRTF is the Fourier Transform of the HRIR. At the time of writing, the main institutes that are working on measuring HRTFs includes the CIPIC International Lab, MIT Media Lab, The Graduate School in Psychoacoustics at the University of Oldenburg, Neurophysiology Lab in University of Wisconsin-Madison, and Ames Lab of NASA.

Precedence Effect

The Precedence Effect, also known as the law of the first wavefront, is a binaural psychoacoustic effect where a single sound is perceived when two identical (or nearly identical) complex sounds are presented in quick succession. The apparent location of the auditory event is dominated by the location of the sound that arrived first (first arriving wavefront), or the louder copy. When a primary sound and a reflection are presented at the same level, summing localization occurs when the delay of the reflection is short (less than 2 ms). With increasing delay, the law of the first wavefront comes into play.

3.1.4 Auditory Spatial Impression

Spatial Impression can be described as the phenomena of how a subject perceives the size and type of a given space based on the acoustics of a presented sound field (actual or simulated). It is an attempt to group all the sensations related to the spatial aspect of the perceived sound. A distinct factor that contributes to the spatial impression is the characteristic "temporal slurring of auditory events" [16], which results from late reflections and reverberations. An auditory effect called *spaciousness* is the characteristic of an auditory event's spatial spreading, and can be divided into two separate, and distinct components [19, 46]. For an alternate classification of spatial impressions, please refer to [35, 65].

1. **Apparent Source Width (ASW)** Size is an inherent attribute to sound sources. When we listen to sounds in the real world, it is easy to estimate the size of a sound event. For example, we understand that the size of the auditory image from an airplane is much greater than the sound of a bee. The perception of the sound's size is not only dependent on the loudness of the event, rather it is a combination of frequency, loudness, and signal duration [17, 50, 49]. The wavelength of a low pitched sound needs a greater distance to unfold, resulting in a larger apparent width compared to high pitched sounds. The distance of an auditory event is typically closely linked to its loudness, which also affects the perception of width. As a result, the

perception of source width decreases over distance, similar to how an object appears visually smaller with increasing distance [51].

The difference and similarities between the two signals entering both ears plays an important role in the apparent source width. Various techniques based on decorrelation of sound sources have been used to change the source width of a sound [34, 38, 39, 19].

2. **Listener Envelopment (LEV)** When a sound is played in an enclosed space, a listener receives 3 instances of the sound: the direct sound, early reflections, and the reverberant sound [11]. The direct sound is primarily perceived right after the sound is played (less than 35 ms), followed by the early reflections created by the walls, ceiling, and physical attributes of the space. The reverberant sound is all the sound that arrives at a listener's ear 80 ms to 100 ms after the initial direct sound is heard. Listener Envelopment can be defined as the degree to which the reverberant sound seems to surround the listener, i.e., to come from all directions [12]. In acoustically treated halls such as the Walt Disney Concert Hall, Boston Symphony Hall, and the Hamburg Elbphilharmonie, sound waves are free to roam around, giving the listener a sense of being enveloped by the sound. Measurements and calculation of Listener Envelopment can be found in [12].

3.1.5 Multimodality

The visual and auditory modalities interact with each other in specifying the nature of an event in an environment. The two modalities receive stimuli in the same surrounding, and often times, an event that is of interest produces a change in both visual and auditory stimuli. For example, faces move as people talk, the sound of ocean waves is correlated with the motion of crashing waves, and the sound of thunder is coupled with lightning. Both these senses work together to answer questions such as "what" and "where", and it seems that these correlations appear either at birth or shortly afterwards [18].

Humans are not only sensitive to the correlation between the two senses, but we also use each sense to correct the analysis of a given scene. In certain cases, the interaction between the two senses can create an interference, forcing a person to either incorrectly perceive the sounds, or visuals [18]. Other research shows that the localization of an auditory event is greatly affected by a visual event, and vice versa [64, 16]. For example, the ability of a ventriloquist to trick the audience into believing that the sound is emanating from a dummy's mouth, instead of the performer's mouth.

Furthermore, we use the correlation of events to determine what the events are. In the case of lip reading, we sometimes determine the content of the speech itself not only from the sound that we hear, but also from the lip movement of the

person speaking, especially in the case of deaf individuals, or in a crowded space. The perceived sound can also be altered by changing the visual event, without modifying the sound source [18].

Auditory space is often mistakenly considered to be analogous to visual space. Although the two share some similarities, they are quite different, and less accurate in some respects.

The auditory system misses some events because the reflected sound mixes with the direct sound, and obscures some properties of the original source, for example in the case of auditory masking [18, 59]. However, sound can bend around large obstructions, while light is unable to, allowing us to perceive events all around us at all times, and not only where the head is facing. The best strategy is to combine the information from the two senses.

3.2 Spatial Sound in Music

> *Spatial separation clarifies the texture; this is particularly important if the music consists of several different layers located in the same pitch register. Spatial separation is equivalent to the separation by register or timbre. That is, just as one can hear separately layers of music that are located in different registers, one can also differentiate layers that originate from different points in space. Spatial separation facilitates greater complexity in the music; more unrelated elements can be heard simultaneously.* (Brant 1967, quoted in Roads 2015)

The use of space as a compositional element can be traced back to the 1500s, for example, in Willaert's works for two spatiallly separated organs and choirs at the Basilica San Marco in Venice [60]. W. A. Mozart (K. 239 and K. 286), Hector Berlioz (Requiem, 1837) and Gustav Mahler (Symphony No. 2, 1895) have all written for spatially separated multiple orchestras and choruses– the tradition known as *polychoral music*. In 1950, Henry Brant began writing instrumental music that incorporated space as an essential element. *Antiphony I* (1953) is one out of hundreds of his works that required a different spatial configuration. Other composers such as Stockhausen and Xenakis occasionally used spatially separated instrumentalists in their works.

Recording devices allowed us to capture the spatial environment. The recording process can be thought of as a double convolution, that is; the convolution of the sound event with the impulse response of the physical space, and further convolved with the impulse response of the recording equipment [61]. Conversely, the projection of sonic energy was enabled by the invention of the loudspeaker, and the idea of spatial composition emerged in the late 1950s. The potential was further explored through the creation of electroacoustic technology, such as amplifiers and tape recorders [60].

Artificial reverberation were commonly used throughout 1930-1970. Echo chambers[5] were used in 1931 at the Abbey Road Studios, while spring reverberators were used in Hammond organs and guitar amplifiers. Plate reverberators were used in the WDR Studio in Cologne, and the Columbia-Princeton Electronic Music Studio.

Figure 3.3: Pierre Henry controlling the Potentiomtre D'espace (Space Potentiometer) in a concert at the Salle de L'Ancien Conservatoire, Paris. From Schaeffer (1952).

The early 1950s saw the rise of spatialization in Musique Concrète. The first concert in 1950 employed multiple turntables mixed in real time. In 1951, Pierre

[5]Sounds played into a physical space using a loudspeaker, and the sounds of the space is picked up by a microphone

Schaeffer and his colleagues used magnetic tape for concert playback, where a *space potentiometer*[6] (Figure 3.3) was used for live spatialization [60]. Throughout the 1950s, John Cage created a number of installations involving multiple sources of sound, and compositions that featured multiple outputs, for example, *Imaginary Landscape No. 4*, *Williams Mix*, and *Variations IV* along with David Tudor. Stockhausen's *Gesang Der Junglinge* was spatialized using five loudspeakers in the West German Radio auditorium.

From 1957 to 1959, the Vertex Concerts were held in San Francisco (including one at the Brussels World's Fair). Presented in a domed theater with special projectors and a 38 channel sound system, the concerts featured visual works by Jordan Belson and music by Stockhausen, Ussachevsky, Takemitsu, and Berio. The movement of sound was controlled using a custom made rotary console, while playback and spatial switching system is based on a 35 mm sprocketed magnetic tape.

The Philips Pavilion (Figure 3.4) in Brussels, designed by Xenakis for Le Corbusier at the Brussels World's Fair (1958) was a structure that contained 400 loudspeakers, controlled via an 11-channel sound system. As stated by Varèse, the sounds were spatially distributed using "the very complex electronic [spatializer] device" where a switching system would allow the sounds to travel based on

[6]Four metal hoops manipulated by a sound projectionist distributed soundtrack to any of the four loudspeakers

Figure 3.4: View, looking upward, of the inside of the Philips Pavilion. The objects on the surface are high-frequency loudspeakers in clusters, the patterns designed by Xenakis. Low-frequency loudspeakers were installed on the ground. From Treib (1996).

programmed "sound routes" during the performance. Edgard Varèse's *Poeme Electronique* (accompanied by a film of images chosen by Le Corbusier), and Iannis Xenakis' *Concret PH* were both realized during this event.

Artificial reverberation algorithms were first developed by Dr. Manfred R. Schroeder in 1961 at the Bell Telephone Laboratories. He used combinations of multiple time delays, filters, and multi-tap delay lines to simulate sound scattering

in space. These systems did not run in real-time when they were first developed, but modern implementations of his model such as the Lexicon 300L and PCM96 can be used in real-time.

The Audium Theater, conceived in the 1960s by Stanley Shaff and Douglas McEachern, is a space for choreographing sounds via a 176 speaker array. Luc Ferrari used the field recorder in his soundscape music works (*Presque Rein*) to capture the natural spatial environment. He then processed and layered multiple (recorded) environments to create amazing fictional landscapes.

EXPO 70 was held in Osaka in 1970, and featured three separate major spatial sound systems. The German pavilion featured 55 Siemens loudspeakers, distributed in seven rings on the interior surface of a geodesic dome, and played Stockhausen's music for 183 days. In the Japanese Steel pavilion, Xenakis performed his 12-channel composition (*Hibiki Hana Ma*) using a 800 loudspeaker system. Experiments in Art and Technology (EAT) curated a project in the Pepsi Cola pavilion, where a dome with 37 loudspeakers were used to spatialize up to 32 sound sources [60].

Salvatore Martirano (1971-1972) and a team of engineers at the University of Illinois built an interactive device called Sal-Mar that controlled an analog synthesizer and distributed the sounds using 250 Poly-Planar (styrofoam) loudspeakers.

Xenakis' sound-and-light spectacle *Polytope de Cluny* was projected in the ancient Cluny Museum in Paris over a 12 channel sound system. The piece ran for 16 months, and was experienced by over 200,000 people.

In 1973, Christian Clozier and his colleagues at the Groupe de Musique Experimantale de Bourges (GMEB) developed a system that projects sounds using loudspeakers that are placed on stage, and within the audience. The first implementation of this system was the Gmebaphone, which allowed the spatial projection to be performed manually by composers.

In the late 1970s, Michel Redolfi was sponsored by UC San Diego's Center for Music Experiment and the Scripps Institute of Oceanography with a goal of broadcasting music underwater. In 1981, he presented *Sonic Waters in the Pacific*, which was the first underwater large-scale concert.

The use of computer algorithms as a means of spatializing sound sources has been ongoing since 1971. Edward Kobrin's HYBRID synthesizer consisted of a digital computer controlling an analog synthesizer, and distributed sounds to 16 loudspeakers. John Chowning in 1971 developed the first software for spatialization with Doppler shift, coupled with a digital reverberator based on Schroeder's design. In 1972, Chowning presented his work *Turenas*[7] via a 360° loudspeaker configuration.

[7]Frequency modulation (FM) synthesis in audio was first introduced via *Turenas*

In 1987, a computer-based 32 channel sound distribution system called *Trails* was developed at Luciano Berio's Tempo Reale Studio in Florence. The development of computer-based spatialization systems have since advanced. These systems include Halaphon (Freiburg), GRAME's Sinfonie (Lyon), the BEAST (Birmingham), Simon Fraser University's AudioBox, the Recombinant Media Lab's Cinechamber (San Francisco), ZKM Klangdom (Karlsruhe), CREATEophone (Santa Barbara), and the AlloSphere (Santa Barbara).

3.3 Spatialization in Electronic Music

> *In the early years of electronic music, the spatial aspect of most compositions was fixed for the entire piece. This was a macrospatial perspective. As composers discovered new techniques, their spatial aesthetic became more refined.* (Roads 2001)

Sound spatialization can be divided into two complimentary parts. The virtual space allows one to spatialize sounds using algorithms, processes, and transformations such as tape echo feedback, electronic delays, spectral filters, phase shift (for source width control), convolution, granulation, panning, and reverberation [60]. On the other hand, the physical space of a concert hall enables one to project sound through multichannel or pluriphonic sound systems.

3.3.1 Virtual Spaces

Morphology of space in the virtual world is malleable. Any space can be transformed within any time period, and the evolution of spaces can be the central focus of a composition [67, 60]. Physically impossible spaces such as continuously changing echo patterns, or the simultaneous presence of different qualities of ambience are able to be realized via virtual spaces. Cross-synthesis (e.g., convolution) allows us to impose the spatial characteristics of a given space on to another sound, creating the illusion that the sound existed in the original space. Using space as an element adds another dimension to the work, analogous to the temporal and frequency dimensions. When the spatial dimension is not used in electronic works, the work suffers from "spatial sameness" [75, 59].

The control of moving and stationary sounds, and the relationships between them (spatial counterpoint) is used to create background, and foreground elements in a sound scene [59, 60, 64]. Contrasts in proximity, from extremely close to extremely far can be used as a compositional methodology. Opposition of sounds can also be achieved by contrasting extremely close sound objects, against deeply reverberated parts. The control for distance of sounds can be achieved using several means, for example via close proximity of the microphone in the recording process. As a post-production technique, the sounds can be manipulated by increasing the amplitude, reverberation, or by applying "presence" filters to boost

the low-frequency range. Headphones allow us to perceive a more "intimate" sound image, as opposed to a typical loudspeaker that is placed several meters away[8].

Spatial depth can also be achieved by a sound's virtual acoustic properties. For example, a sound that is sent through a low-pass filter, and reverberation will subside into the background, while sounds that contain high frequencies will appear in the foreground [62]. Background texture can also be achieved by an omnipresent, unobstrusive, or repeating sound element, which functions as a structure for the piece. An example of this can be seen in the use of a "low-level granulose background texture" in Horacio Vaggione's *Nodal* and *Agon* [60] (Figure 3.5).

Microphone techniques and spatial processing can be thought of as analogous to the use of camera angle, lens perspective, and depth of field. Consequently, we see more composers leaning towards cinematic use of space as a compositional element, for example in Luc Ferrari's *Presque rien no. 1* and Jean-Claude Risset's *Sud*.

The microsound realm deals with the smallest perceivable structure of sound, and allows one to control the spatial properties in a precise manner, such as assigning an independent spatial position for every grain. While it is true that the grains can share a spatial position, assigning a unique position for each grain cre-

[8]Wave Field Synthesis is a specific case that enables the sounds to appear "out of" the loudspeaker, around the listening field

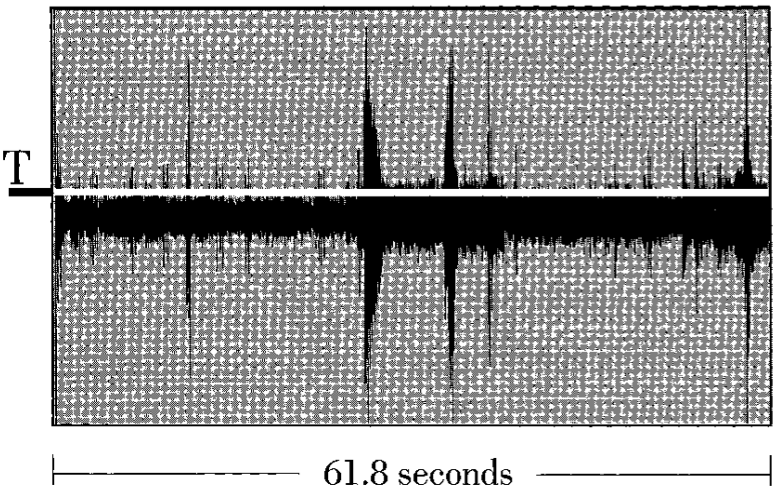


Figure 3.5: The amplitude envelope of the first 61.8 seconds of Agon by Horaciio Vaggione. The line marked T indicates the amplitude threshold between the foreground peaks and the background granulations. From Roads (2001).

ates a three-dimensional sound image. This can be achieved by manually placing each grain in space via a sound editor or mixing program, or via automatic scattering algorithms that assign a position for each grain in virtual space. Examples of high level control for the algorithms include fixed spatial positions, panoramic motion from one position to another over time, random spatial positions, or density of grains per spatial position. Depth of the spatial image can be controlled by adding selective reverberation to each grain, for example based on probability functions. This effect is most effective at low densities, as it fuses into a continuous background reverberation at high densities.

Modulation with particles begin with the process of generating patterns of synthetic grains, and distributing them across N channels. The amplitude envelope of each grain is extracted, and imposed (time domain multiplication) on to another source signal. Convolution [32, 59] with granular clouds is a powerful tool for sound spatialization. An asynchronous cloud of grains can be thought of as an Impulse Response of an unusual virtual environment. If the signal used is a brief source, a sparse cloud is generated, i.e., a statistical distribution of echoes based on the temporal arrangement of the asynchronous cloud. As we increase the grain density, the echoes start to fuse into an irregular quasi-reverberation effect [59]. Time smearing effects occur when a source with a sharp attack is used, whereby each grain generates an echo of the attack. The time smearing is smoothed out into a strange colored reverb (color determined by spectrum of grains) if a source with a smooth attack is used. For low density synchronous clouds, convolution results in metrical rhythms resembling tape echo. As we increase the density, the echoes fuse into buzzing, ringing, or rippling sonorities [59], where the identity of the source may be destroyed. For a detailed analysis on the effects of cloud amplitude envelope, particle envelope, particle duration, particle frequency, and window type in sectioned convolution, please refer to *Microsound*, pp 226-231.

Tape Echo Feedback is a classic transformation technique that creates a ping-pong panning echo effect [60]. The classic version of this technique consist of

Figure 3.6: Curtis Roads and the author controlling Tape Echo Feedback in real time

an analog tape recorder with continuous variable speed control (first developed by Werner Meyer-Eppler), and was featured in Stockhausen's *Kontakte*. Sounds are played into a tape recorder, and immediately fed back into the tape recorder. Optionally, the feedback signal can be passed through a filter and mixed with new incoming sounds.

Various effects can be generated from this technique, based on the parameter settings. At low level feedback signal, we hear a series of faint echoes, while a moderate to high feedback signal creates echoes that may be louder than the

incoming signal, or may potentially become high self-sustaining oscillation. The pitch and echo rate can be changed via varispeed control, while the spectrum of the feedback can be altered using a filter.

This technique is inherently unstable, where self-sustaining feedback loops can occur at any time. Therefore, it requires two people to apply careful control of feedback levels, tape speed, and filter settings in real time (Figure 3.6). In musical applications, the goal is to "surf on the edges of self-oscillation"– a zone of morphosis, while also applying damping forces to the feedback process through spectrum shaping and control of the feedback amount [60].

Audio example 3.1. Result of Tape Echo Feedback.

Artificial reverberators are commonly used in electronic music. This technique includes Schroeder's model (Section 3.2), physical models of simulated spaces such as *geometric models* (beam and ray tracing), *waveguide networks*, and *feedback delay networks*. Convolution, as described above, can also be seen as a form of reverberator (Figure 3.7, as it imposes the spatial characteristic of a recorded space[9] on to a different sound. It is possible to assign a different virtual space for each sound object, or grain.

[9]The Impulse Response (IR) can either be recorded, or downloaded from extensive libraries. The IRs can also be synthetically created in order to model strange, imaginary spaces

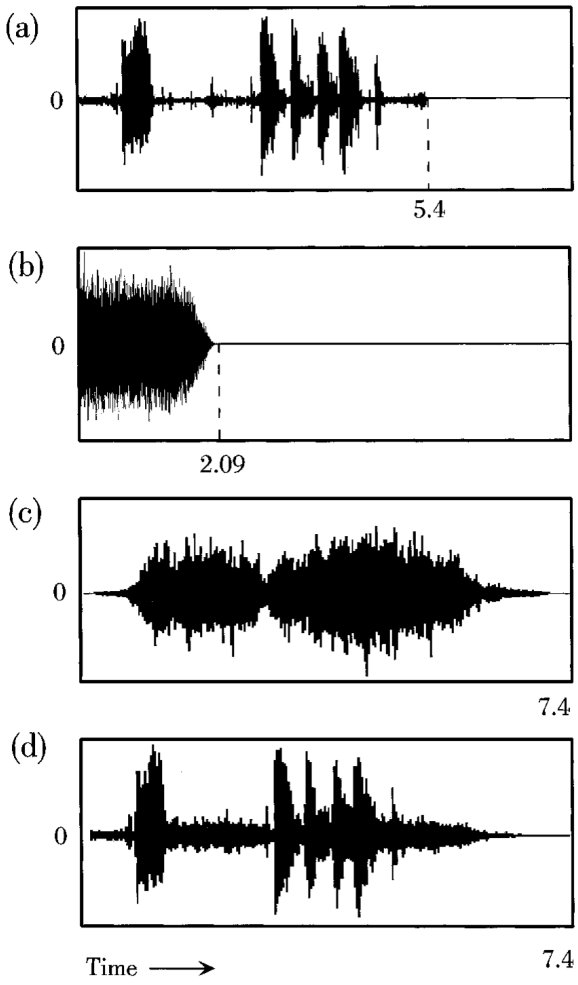


Figure 3.7: Reverberation by granular convolution. (a) Speech input: "Moi, Alpha Soixante." (b) Granular impulse response, consisting of one thousand 9-ms sinusoidal grains centered at 14,000 Hz, with a bandwidth of 5000 Hz. (c) Convolution of (a) and (b). (d) Mixture of (a) and (c) in a proportion of 5 : 1, creating reverberation around the speech. From Roads (2001).

So far, the reverberation is treated as a consequence of a sound event. Alternatively, the reverberation can be treated independently from the original sound event. For example, one can input a sound source to a real-time reverberator, and record the reverberation on a separate track. The reverberation track can then be pitch-shifted, and stacked together (with a different transposition) to create a *reverberation chord* [60].

3.3.2 Physical Spaces

Sound systems in a physical architecture and virtual acoustics work hand in hand to present a spatial impression, creating an interplay between the static architecture of the hall, and the dynamic virtual acoustics of the music. Sound projection systems can be divided into 3 categories based on their loudspeaker configuration, namely:

1. **Stereophony, quadrophony, octophony:** Lateral array in front and around the listener

2. **Periphony:** The configuration above extended to vertical dimension [33]

3. **Pluriphony:** Orchestra of loudspeakers on stage– spatial diffusion of electronic music, operated using a sound mixing console for spatial projection [22]. Example of this system includes the GMEBaphone/ Cybernèphone

An electronic piece embodies its own virtual space, as determined by the composer/ performer on a specific playback system. Playing back the piece in a different physical setup (e.g., headphones, or different number of speakers and configuration) creates a different spatial impression. The piece can be adapted to the loudspeaker configuration of the hall, i.e., optimizing the playback based on the venue, or reinterpreted to enhance and extend the virtual spaces that exist on the playback medium.

Different aspects of the venue's physical architecture *colors* the virtual sound. Different spaces resonates at different frequencies based on the size and geometry of the space. Room reflections happen when the projected sound energy hits various surfaces– some energy is reflected, some absorbed, some transmitted through the surface. It is also dependent on the number of spectators and their attire [60, 25]. Every space has its own unique sound and introduces reverberation, based on the pattern of sound reflections. A good space diffuses the energy by creating random reflections, and scatters reverberant energy equally to all areas[10]. In contrast, a poor quality of a concert hall is one that creates focused reflection patterns, resulting in echoes and uneven resonances. Ancillary vibration is an

[10]Herzog and De Meuron's Elbphilharmonie (Hamburg, Germany) makes use of algorithms to generate a unique shape for each of the 10,000 *gypsum fiber acoustic panels* that line the auditorium's walls to create unique absorbtion and scattering patterns

artifact caused by the buzzing and rattling sounds created by various objects in the space, such as light fixtures and cable trays.

Loudspeaker type and configuration is arguably the most important factor to consider, as it determines the reproducibility of the sounds. Loudspeakers can be measured based on its bandwidth, and distortion. The selection is also dependent on the size, weight, power requirements, cost, and ultimately, the listener's subjective preference. Subwoofers enhance the lower ends of the spectrum, giving a more pronounced definition of bass frequencies. However, a disadvantage is that the bass seems detached from the rest of the sound, especially when the sound is moving in space. The placement of loudspeakers in space, and the number of loudspeakers is based on the spatialization algorithm used, and physical constraints [7]. The concert halls listed at the end of Section 3.2 are a few examples of loudspeaker configurations, but the topic is currently a research area that is active with scientific and artistic experimentation.

Pluriphonic Sound Projection

Live diffusion in a pluriphonic sound setup is a technique that unifies the virtual space and the physical space of a concert hall [60]. It is typically achieved through the use of an interface that maps input channels to loudspeakers in the hall. Real-time spatialization enhances and extends the virtual space, and allows

one to create a different virtual movement. Horacio Vaggione states that "The goal is not to maintain a stereotypical stereo image, but rather to break it to better reconstitute the plurality contained inside the work." [74, 60].

This technique allows each listener to experience a unique spatial perspective [23]. Instead of a downmix, where many input channels are reduced to a small number of outputs, a small number of inputs are distributed to many outputs– a process known as *upmixing*. Projected sounds can appear from above, below, and within the audience.

This process is made possible by the installation of possibly dozens of loudspeakers around, and even within the audience, making it possible to continuously move sounds around the architectural space. Consequently, it creates a sonic experience that can be appreciated from different points of view, each evoking different impressions, such as dimensional, and immersive attributes. Roads [60, 59] lists three principles concerning the art of pluriphonic projection:

1. The experience of an electronic music composition in stereo format can be greatly enhanced by a spatial performance in concert, whether semi-automated or diffused by a musician in real-time. Alternatively, the spatialization of a multichannel composition can be realized in the studio and mapped to a multichannel sound system in the concert venue.

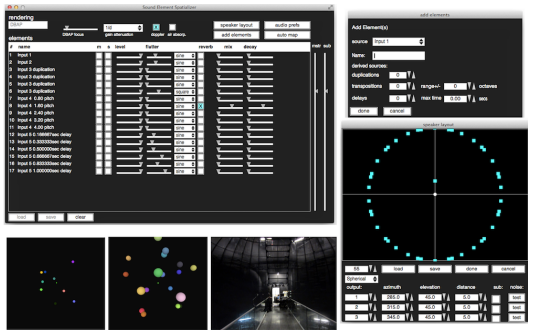
2. The sound projection system can offer a variety of contrasting spatial images through the arrangement of multiple loudspeakers around the audience, across the front stage, above, within, and below the audience. Thus each listener has a unique perspective, and there is not necessarily a "correct" position from which to hear the music. In performance, the composer selects particular spatial images to highlight certain aspects of his or her work and choreographs transitions from scene to scene. Deploying multiple loudspeakers onstage makes it possible to project a sound image rivaling the complexity of an orchestra.

3. While a single type of loudspeaker guarantees uniformity in sound quality, it is also possible to mix different types of loudspeakers in the same pluriphonic system, with the most common case being full-range versus subwoofers. Each type offers a particular voicing that may be useful in articulating a specific musical texture.

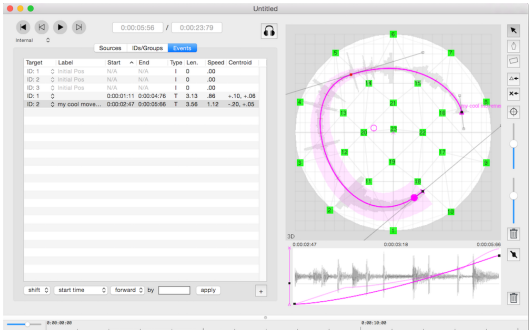
The end of Section 3.2 lists a number of pluriphonic systems that use dozens and hundreds of loudspeakers as part of their configuration. As a consequence to these developments, direct hands on control of each source becomes more complicated, if not impossible. Spatial gestures such as rotation of multiple sources at different speeds and angles, are too complicated to perform in real time.

Solutions to the problem include non-real-time scripting or sequencing where the composer pre-spatializes the channels in the studio. This method ensures that the piece can be played in various venues with minimal requirements. Real-time interactive diffusion systems allow one to control the upmixing via semi-automatic high-level controls. These systems make use of algorithms to control low-level details, and allow musicians to control large spatial gestures. Examples of such systems include UCSB's *Sound Element Spatializer* (Figure 3.8 (a)), ZKM's *Zirkonium* (Figure 3.8 (b)), IRCAM's *Le Spatialisateur* (Figure 3.8 (c)), and TU Berlin's *The SoundScape Renderer* (Figure 3.8 (d)). Roads' most important strategy in projecting sound from a stereo source to a pluriphonic system is using *Spatial Chords* [59, 60, 21].

The classic means of upmixing is to decorrelate two signals by applying phase shift, filter, delay, or granulation [38]. Using a similar technique, Cabrera and Kendall controls the source width of a sound via *Sinusoidal Partial Modulation* [20]. Scatter [44] decomposes a sound into time frequency representations using a technique known as *Dictionary Based Pursuit* [70], and spatializes sound based on features found in the Time-Frequency analysis. On the other hand, basic techniques such as filter banks, phase shifts, and delay lines can be used to separate different aspects of the sound, and spatialized separately. Although the techniques

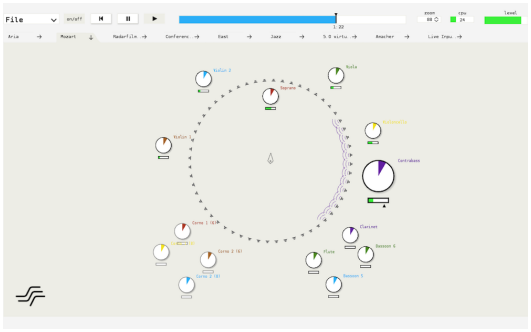


(a) Sound Element Spatializer



(b) Zirkonium

(c) Le Spatialisateur



(d) Soundscape Renderer

Figure 3.8: Graphical User Interface for spatialization systems

listed here functions as a way to diffuse sounds in space, the aesthetic challenge lies in articulation of musical structure through the spatial projection.

3.3.3 Sound Field Synthesis

There are three major approaches to rendering spatial sounds. These techniques combine the loudspeaker configuration (array of loudspeakers surrounding the listener), and spatial signal processing (diffusion of audio signal to loudspeakers). The three techniques are as follows:

1. **Vector Base Amplitude Panning (VBAP)** is an extension of equal power panning to multi-channel configurations. Stereophonic techniques project a signal on to a stereo sound field with a phantom source between two loudspeakers. In contrast, VBAP projects on to three loudspeakers arranged in a triangle configuration, which allows the panning to occur both vertically and horizontally [53]. A typical configuration for VBAP is regularly spaced loudspeakers around and above the audience. The spatial resolution of this technique, or the quality of rendering is greatly dependent on loudspeaker density, and the regularity of loudspeaker placement. Another similar technique is known as Distance Base Amplitude Panning (DBAP) [40]

2. **Wave Field Synthesis (WFS)** is based on the Huygens principle, which states that any wavefront can be regarded as a superposition of elementary spherical waves [13, 14, 76, 8]. The theoretical framework was initially proposed by Snow in 1955, but the development was halted due to technical constraints [68]. WFS does not rely on psychoacoustics, but rather renders the whole sound field corresponding to the scene, resulting in a listening position that is not only specific to one location. It can create the impression of a 3D virtual point source by using a combination of tightly spaced loudspeakers– every loudspeaker contributes to the reproduction of the virtual source.

3. **Ambisonics** is a technique for encoding spatial information using spherical harmonic functions that handles both *surround* (flat multichannel) and *periphony* (multichannel with height) [33]. Ambisonics is originally based on microphone recording (sound-field microphone), but is later available as a synthesis technique for existing audio files. The technique can be divided into two independent processes: encoding and decoding.

   In the encoding stage, the spatial sound field is projected on to spherical harmonics, analogous to how the Fourier Transform projects a signal on to sinusoids of different frequencies. In the decoding stage, all loudspeakers are generally used to localize sound, as opposed to other techniques that

use only adjacent speakers, such as VBAP. One of the great advantages of Ambisonics is that it decouples the encoding and decoding processes, allowing for great scalability. High-order Ambisonics (HOA) corresponds to the number of spherical harmonics used in the encoding process– which is typically larger than four channels. A more detailed discussion on Ambisonics will be provided in Section 4.1.1.

The vast possibilities introduced by immersive sound systems and spatialization techniques have given birth to a genre of music that focuses on the central theme of spatial structure as the main narrative. All the other elements serve as a support structure for the spatial organization. Articulation of space necessitates relationships between elements, such as the play between attracting (similarities) and opposing (contrasts) forces. Lateral position, vertical position, image width, and image depth are examples of various dimensions in the sonic space that plays an important role in constructing this relationship. Table 3.2 [60] lists a basic repertoire of possible spatial oppositions.

Foreground (present)	Background (obscured or reverberated)
Sole position in space	Multiple positions in space (spatial chords) forming geometrical shapes
Sole position in space	Spatial envelopment (sound from all sides)
Panning by related pairs of loudspeakers, e.g., from front left and right to rear left and right	Positioning and panning by arbitrary collections of loudspeakers, creating spatial chords, e.g., from upper front left and lower rear right to lower middle left and upper front right, generalized to n channels
Fixed position in space	Moving position in space
Fixed position in space	Scattered position in space (through granular decorrelation)
Fixed position in space	Oscillation between two positions in space, a kind of "spatial trilling".
Fixed dispersion pattern	Variable dispersion pattern (changes of apparent source width, possibly modeling the dispersion pattern of an acoustic instrument, horn, lens, or other source)
Sources positioned at two extreme poles	Sources filling in a stereo field, including the center
Fixed source geometry	Rotating source geometry
Slow motion	Fast motion
Periodic movement (sinusoidal, pulse, linear pan, exponential pan, logarithmic pan)	Random movement, juxtapositions in space
Spatialization is organized as an independent parameter, apart from pitch, rhythm, timbre, etc.	Spatialization linked to mesostructural musical function in coordination with other parameters; the spatial design helps to articulate structural transitions and "changes scene" on musical phrase boundaries. These changes can be linked with any of the oppositions in this table. For example, a transition from one phrase to another could be tied to a transition from one spatial chord to another
Global spatialization, such as global reverberation	Multiscale spatialization: phrases, objects, microsounds can be all given individual spatial characteristics

Table 3.2: Spatial Oppositions. From Roads (2015).

Spatialization is independent of frequency band and formant structure	Spatialization by spectrum, i.e., applying spatial filters that pan sounds depending on their frequency band (Wenger and Spiegel 2004; Sturm et al. 2008, 2009) or formant (as in pulsar synthesis, see Roads 2001, 2002)
Spatialization is independent of sound duration and amplitude	Spatialization by grain size and/or amplitude (Sturm et al. 2008, 2009)
Linear motions, from loudspeaker to loudspeaker	Coordinated geometric rotations at different speeds and directions (circular, elliptical, Lissajous, etc.)
Unidirectional rotation	Multidirectional rotation, including contrary motion (e.g., two sounds spinning in opposite directions)
Horizontal panning	Vertical panning (above and below the listener)
Circular rotation at a constant rate	Spiral rotation with acceleration
Panning without Doppler shift	Panning with Doppler shift
Spatial movement of multiple sounds with swarming or flocking behavior (sounds loosely follow one another; correlated movements)	Spatial movement of multiple sounds with independent trajectories (uncorrelated movements)
Fixed spatial perspective of virtual sounds	Variations in perspective ("cinematic" use of virtual space so that certain sounds appear to be recorded very closely while others appear to be distant)
Conventional loudspeaker dispersion pattern	Superdirectional sound beams
For fixed position sounds, fixed width of the sound image across multiple loudspeakers	Variations in the width of the image across multiple loudspeakers
Conventional spatial projection bounded on its inner surface by a perimeter of loudspeakers	Wavefield synthesis in which the sound emerges from the loudspeaker perimeter and comes into the room
Multichannel spatial image	"Collapsed" spatial image to a single point or to an overall monaural image

Table 3.3: Spatial Oppositions (Continued)

Chapter 4

Spatiotemporal Granulation

Digital Signal Processing has introduced various transformations which allow us to manipulate the temporal and spectral aspects of sound. The first step to many of these techniques involve the automatic segmentation of sonic material into microtemporal particles [59].

As an extension, the first step to spatial transformations of microsound is the automatic segmentation of (spatially encoded) sonic material into *microspatiotemporal particles*.

4.1 Theory

The classical method of granulation captures two perceptual dimensions: time-domain information (e.g., starting time, duration, envelope shape), and frequency-domain information (e.g., the pitch of the waveform within the grain and the spectrum of the grain) [59].

The method described in this document granulates space, and adds another dimension to this representation: spatial-domain information. The fundamental premise of this method lies in the segmentation of spatial sound information into grains which are localized in space and time. These grains will henceforth be individually referred to as a *Spatiotemporal grain* (Figure 4.1).

Once the spatial information is decomposed into individual spatiotemporal grains, various manipulations could then be applied to transform the original sound field, described in Section 4.3.

4.1.1 Encoding Spatial Sound

There are several microphone technologies that allow the capturing of spatial information, such as *X-Y/ Blumlein Pair*, *Decca Tree*, and *Optimum Cardioid Triangle*. However, these technologies do not capture the complete full-sphere information of spatial sound.

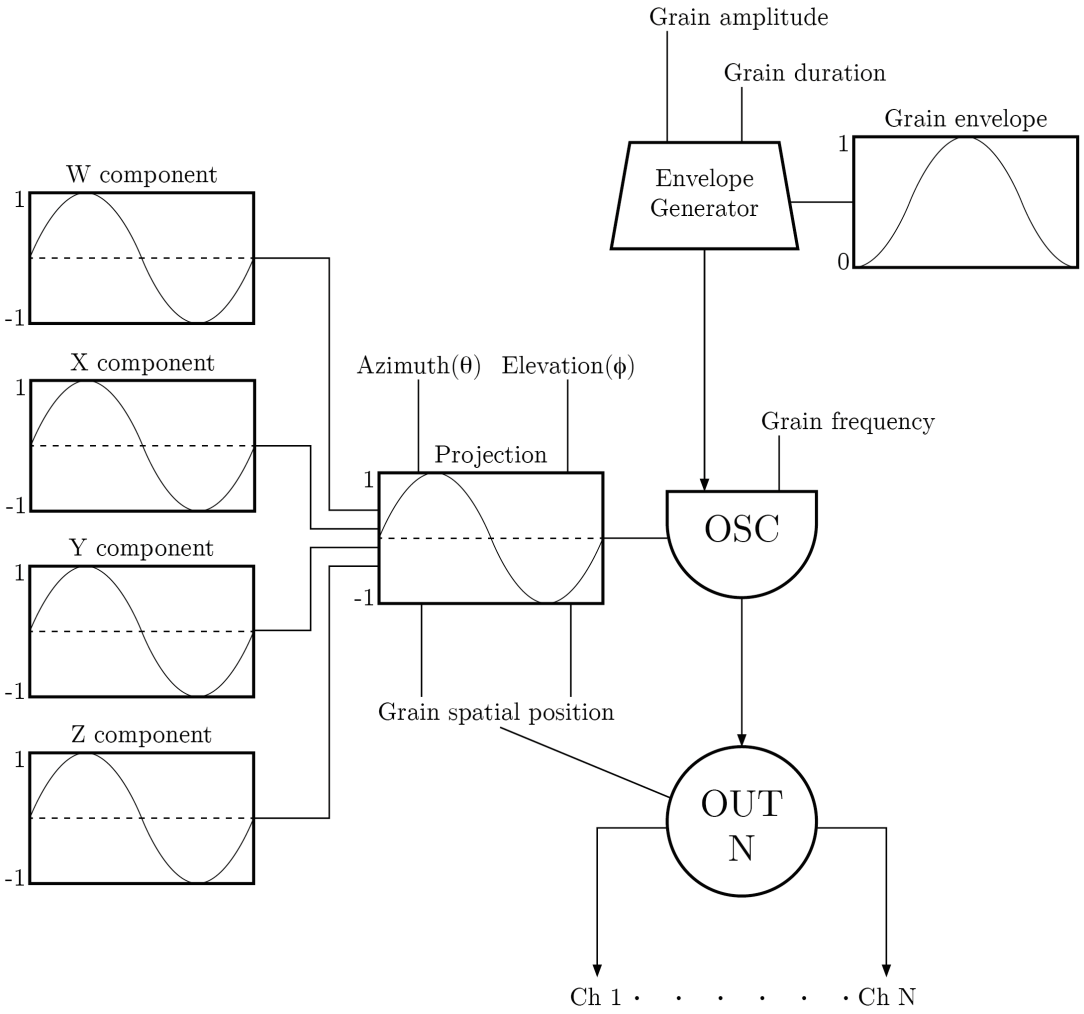


Figure 4.1: Block diagram of a basic spatiotemporal grain generator

On the other hand, *Ambisonics* is a technique that captures periphonic spatial information via microphone arrays[1], such as the "SoundField Microphone" [33]. It is important to note that using this technique, sounds from any direction are

[1]In addition to recording using microphones, Ambisonic signals can also be virtually encoded

treated equally, as opposed to other techniques that assume the frontal information to be the main source, and other directional information as ambient sources.

The spatial soundfield representation of Ambisonics is captured via *Spherical Harmonics* [33]. Spatial resolution is primarily dependent on the order of the Ambisonics signal, i.e., order of Spherical Harmonics (Figure 4.2).

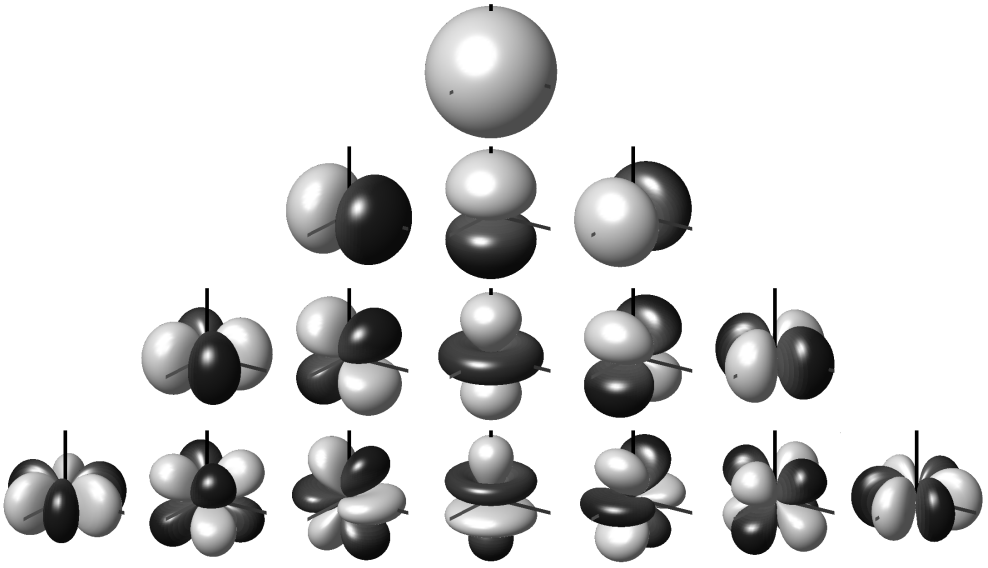


Figure 4.2: Spherical Harmonics up to degree 3, as used in third-order Ambisonics [6]

A first-order encoded signal is composed of the sound pressure W(Eq. 4.1), and the three components of the pressure gradient X(Eq. 4.2), Y(Eq. 4.3), and Z(Eq. 4.4), representing the acoustic particle velocity. Together, these approximate the sound field on a sphere around the microphone array.

$$W = \frac{1}{k}\sum_{i=1}^{k} S_i(\frac{1}{\sqrt{2}}) \tag{4.1}$$

$$X = \frac{1}{k}\sum_{i=1}^{k} S_i(\cos\phi_i \cos\theta_i) \tag{4.2}$$

$$Y = \frac{1}{k}\sum_{i=1}^{k} S_i(\sin\phi_i \cos\theta_i) \tag{4.3}$$

$$Z = \frac{1}{k}\sum_{i=1}^{k} S_i(\sin\theta_i) \tag{4.4}$$

4.1.2 Decoding Spatial Sound

One of the strengths of Ambisonics is the decoupling of encoding, and decoding processes. This allows the captured sound field to be represented using any type of speaker configuration.

In practice, a decoder projects the Spherical Harmonics on to a specific vector, denoted by the position of each loudspeaker θ_j. The reproduction of a sound field without height (surround sound), can be achieved via Eq. 4.5.

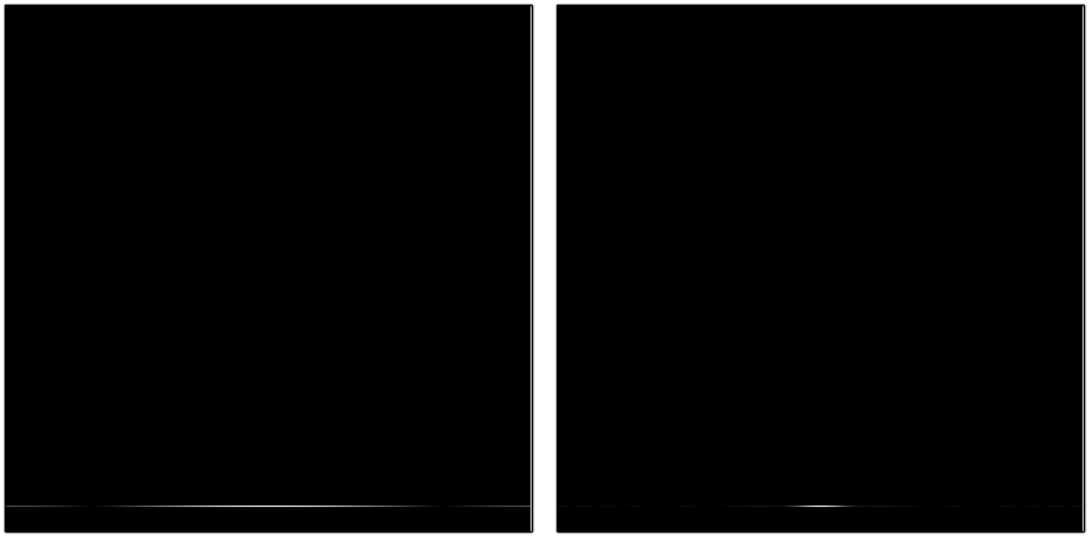
$$P_j = W(\frac{1}{\sqrt{2}}) + X(\cos(\theta_j)) + Y(\sin(\theta_j)) \tag{4.5}$$

4.2 Analysis

4.2.1 Spherical Harmonics Projection

Consider the case where we have N number of loudspeakers arranged on the lateral plane– in a circle without height. In the case where N is 360, each speaker is essentially playing back the sounds that originated from the captured sound field at 1° difference. Instead of playing the sounds from 360 loudspeakers, we can use the information as a means to specify different sounds from different locations. This process results in the segmentation of space by spatially sampling the spherical harmonics.

This forms the basis for extracting the spatiotemporal grains from spatially encoded sounds. Although the resolution of spatial segmentation can be increased, the maximum resolution is limited by a few factors outlined in Section 3.1.2 and Section 4.2.4. Figure 4.3 shows the difference in spatial resolution for a decomposed sound field. The two Ambisonic signals contain synthesized 440 Hz sine tone, encoded using first order, and eleventh order respectively. The results prove that the decomposition of a higher order Ambisonic signal results in a higher spatial resolution. Correspondingly, the uniqueness of the resulting spatiotemporal grains is greater.

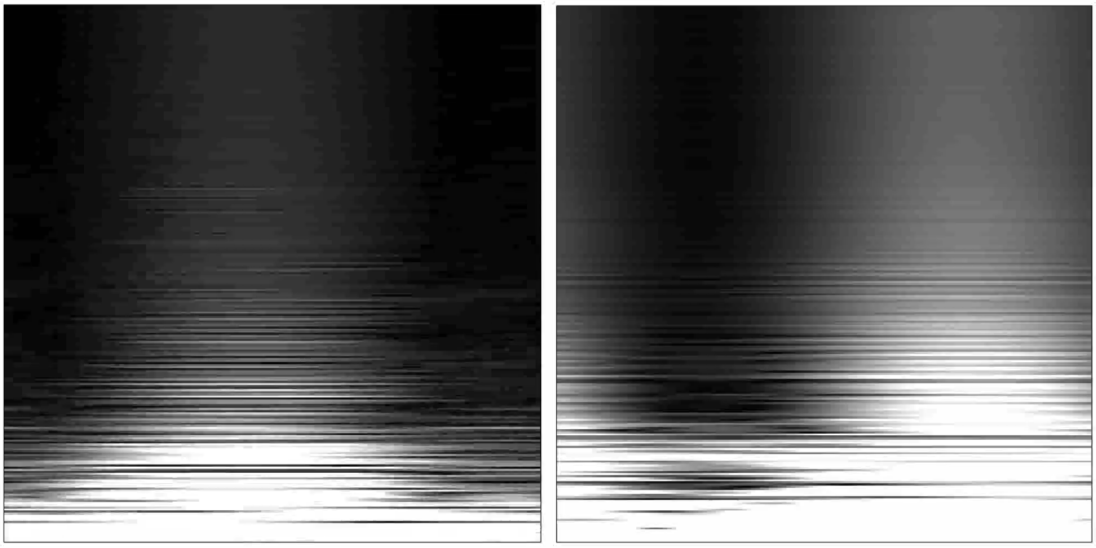


(a) First order (b) Eleventh order

Figure 4.3: Varying order for a 440Hz sine tone encoded via Ambisonics. X-Axis= Azimuth (0°- 360°), Y-Axis= Frequency bin, Intensity= Magnitude of bin

Figure 4.4 shows the decomposition of signals that were captured using a compact spherical microphone array (SoundField ST350). The X axis corresponds to the direction, from 0° to 360°; the Y axis corresponds to frequency bins from 20 Hz to Nyquist frequency (22050 Hz); and the plot's intensity represents the magnitude of each bin. The analysis window size is 512 samples.

If we were to look at the frequency content of these extracted grains in the same temporal window (Figure 4.4 (a)), we can deduce that each spatially localized grain contains a unique spectrum.



(a) Time (in samples): 39424 (b) Time (in samples): 50688

Figure 4.4: X-Axis= Azimuth (0°- 360°), Y-Axis= Frequency bin, Intensity= Magnitude of bin, Window size = 512 samples

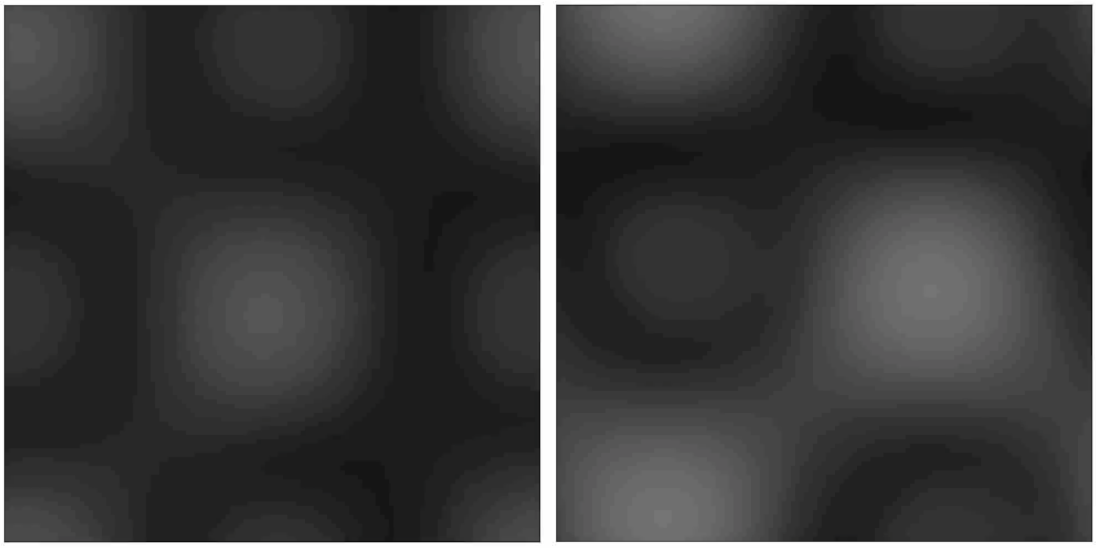
Periphonic projection

Equation 4.5 could also be extended to include height information, i.e., extracting every spatiotemporal grain in space (Eq. 4.6).

$$P_j = W(\frac{1}{\sqrt{2}}) + X(\cos(\theta_j)\cos(\phi_j)) + Y(\sin(\theta_j)\cos(\phi_j)) + Z(\sin(\theta_j)) \tag{4.6}$$

The result of this decomposed sound field can be represented as a 2 dimensional array of spatiotemporal grains, in the same temporal window (Figure 4.5). In this plot, the X axis corresponds to the azimuth, from 0° to 360°; the Y axis

corresponds to the elevation, from 0° to 360°; and the intensity represents the magnitude of each bin. The analysis window size is 512 samples.



(a) Time (in samples): 39424 (b) Time (in samples): 50688

Figure 4.5: X-Axis= Azimuth (0°- 360°), Y-Axis= Elevation (0°- 360°), Intensity= Energy of localized spatiotemporal grain, Window size = 512 samples

Each snapshot of time represents one spatiotemporal frame (Figure 4.5 (a) and (b) respectively). By successively lining up these temporal *slices* (frames), we gain a representation of the full decomposition, i.e., every spatiotemporal grain in space and time.

4.2.2 Spectral Analysis

In the analysis above, we have proven that each decomposed spatiotemporal grain contains a unique spectrum. This information is only the surface of features that could be extracted from the sound field. Figure 4.9 shows the result of performing spatiotemporal granulation on different sound sources. The contrast between each result shows that analysis of each grain's spectrum would provide a wealth of information about the sound field. The possibilities of spatial feature extraction, and descriptors are further discussed in Section 4.5.4.

Based on figure 4.4, we can extract a variety of features, such as the harmonic relationship between each grain, or spectral centroid of each grain. These features could serve as a building block for other higher level information about the content of the signal, or the space. For example, the frequency rolloff allows us to estimate the direction of a particular sound object. This estimation could be based on the relationship between bins in the Fourier Transform, or by simply taking the average magnitude for each grain.

Video example 4.1. Result of detecting source direction by estimating the peak of spectral bins.

4.2.3 Reconstruction of Spatial Sound

Prior to performing any transformations, we need to ensure that the original signal is able to be reconstructed from the decomposed space. The process for re-encoding the signal is the inverse of the decomposition. In other words, each Spatiotemporal Grain is weighted based on it's direction in space, and summed to recreate the original Ambisonic components.

Figure 4.6 and 4.7 shows both the original 4 channel Ambisonic components, and the reconstructed signals. The sound of fireworks was used in the analysis shown in Figure 4.6 while the sound of a choir was used in the analysis shown in Figure 4.7.

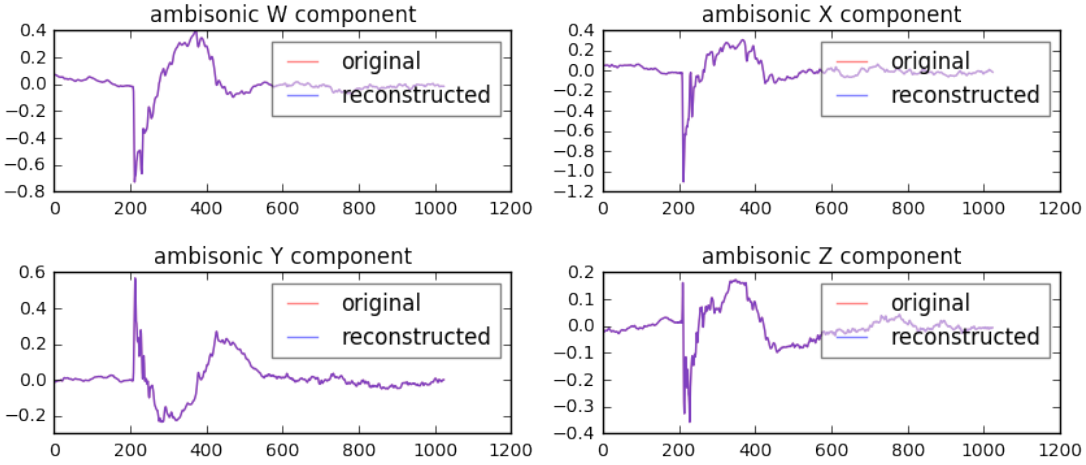


Figure 4.6: Reconstructed Ambisonic components. Source: Fireworks. X-Axis= Time (in samples), Y-Axis= Amplitude, Window size = 1024 samples

Audio example 4.1. B-Format Fireworks.

Audio example 4.2. B-Format Choir.

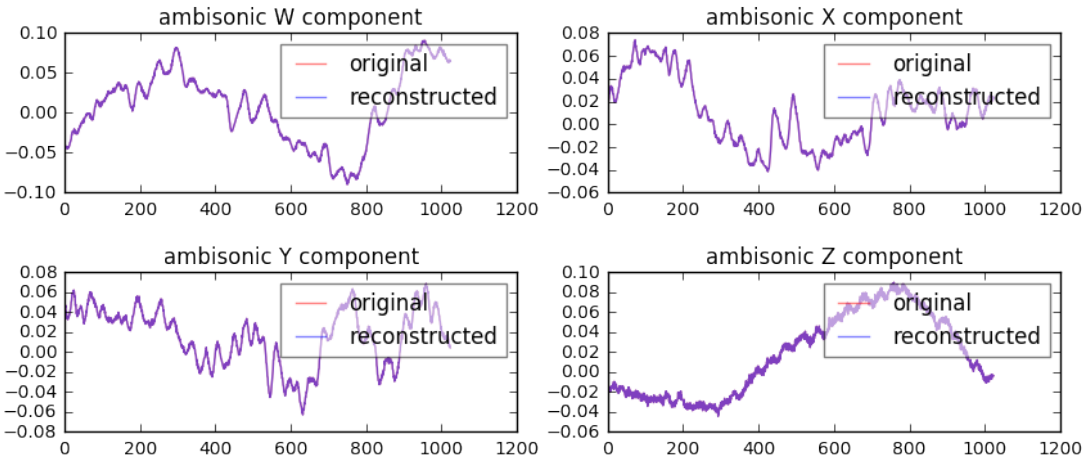


Figure 4.7: Reconstructed Ambisonic components. Source: Choir. X-Axis= Time (in samples), Y-Axis= Amplitude, Window size = 1024 samples

4.2.4 Spatial Resolution

As mentioned in Section 4.1.1, Ambisonic signals can be generated by encoding sonic materials into Higher Order Ambisonics, or by capturing the sound field using compact spherical microphone arrays.

To determine the spatial resolution of Ambisonic signals, sine tones were generated using different orders of Spherical Harmonics. Figure 4.8 shows the decomposition of varying orders of Ambisonic signals through Spatiotemporal Granulation.

The four figures correspond to the decomposition of 1st order, 3rd order, 7th order, and 11th order Ambisonic signals respectively. This experiment confirms that the order of Spherical Harmonics is one of the factors that affect the spatial resolution for Spatiotemporal Granulation.

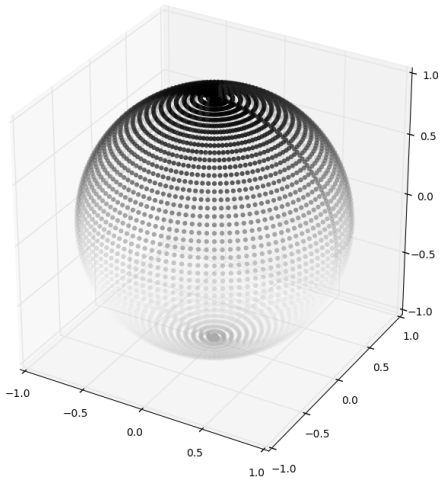
Figure 4.9 shows the result of decomposing four different types of sound sources– fireworks, steamtrain, gulls, and buzzard. The four figures show how the characteristics of the sound event affect the resolution of the decomposition. Transient sounds such as fireworks tend to produce unique spatiotemporal grains, while sustained sound events such as the steamtrain produces highly correlated spatiotemporal grains.

Audio example 4.3. B-Format Steamtrain.

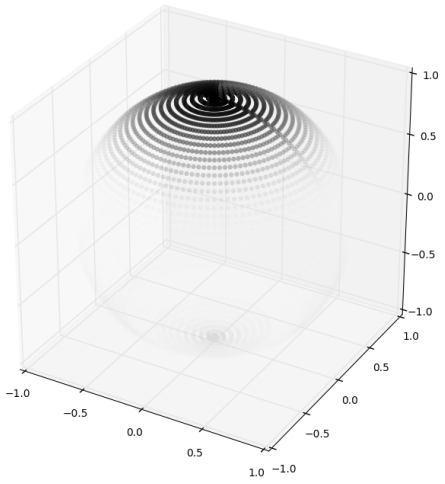
Audio example 4.4. B-Format Gulls.

Audio example 4.5. B-Format Buzzard.

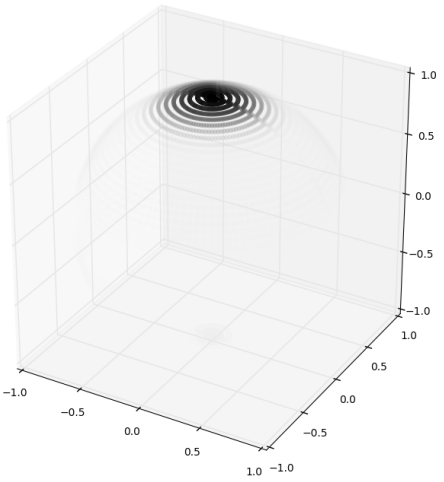
Furthermore, the content of the signal also effects the uniqueness of the spatiotemporal grains. For example, grains that are extracted from the sample of gulls are more similar to one another (based on the frequency rolloff at the direction of sound event), as opposed to the more spread spectrum of grains extracted from the buzzard sample.



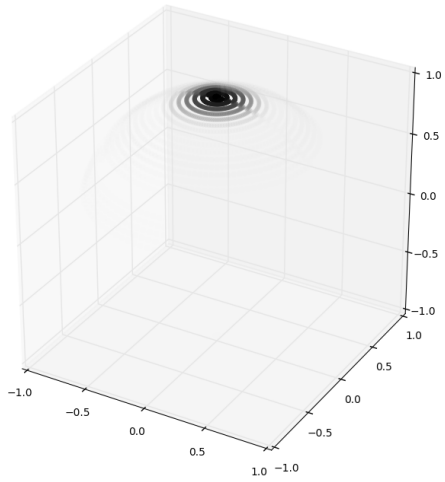
(a) First order



(b) Third order



(c) Seventh order

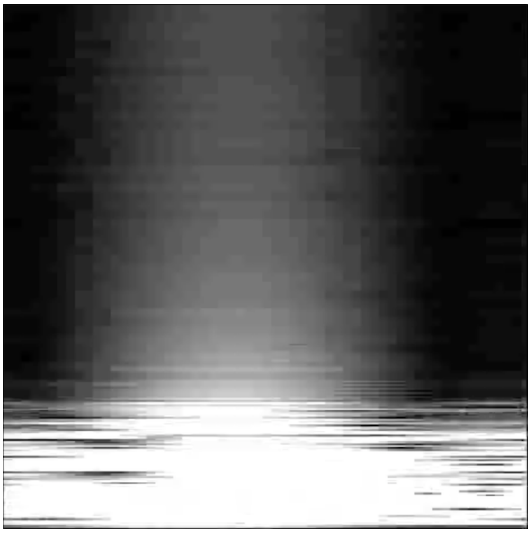


(d) Eleventh order

Figure 4.8: Ambisonics encoded sinusoid of varying orders. Resolution for decomposition: 1° Azimuth, 1° Elevation

(a) Fireworks

(b) Steamtrain



(c) Gulls

(d) Buzzard

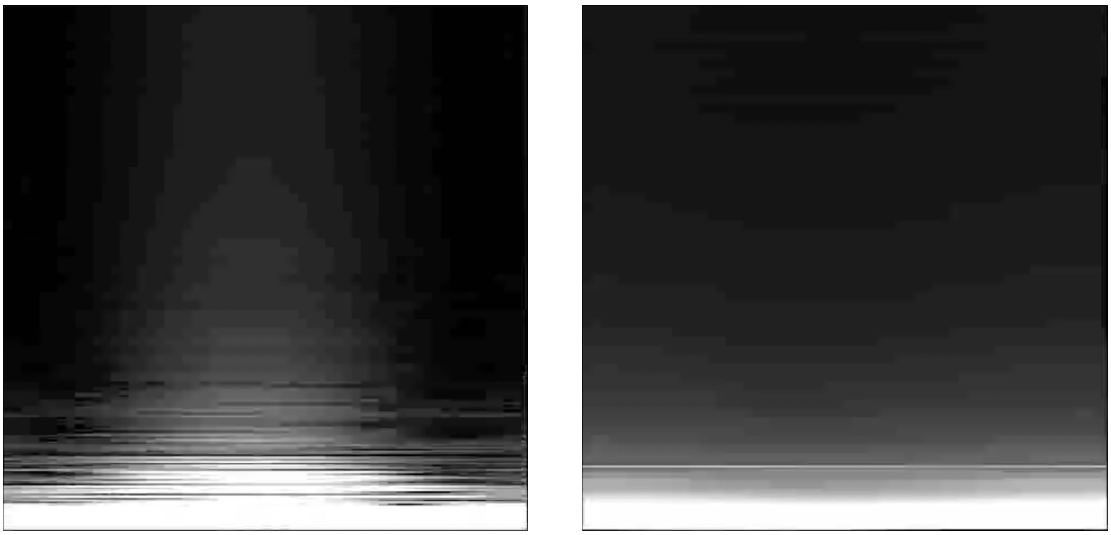
Figure 4.9: Spatiotemporal granulation performed on different types of sources. X-Axis= Azimuth (0°- 360°), Y-Axis= Frequency bin, Intensity= Magnitude of bin, Window size = 512 samples

Another factor that affects the spatial resolution is the physical space where the sound was captured in. For example, grains that are extracted from samples recorded in reverberant halls tend to be less unique to one another. This is caused by the temporal smearing introduced by the hall, as discussed in Section 3.3.2.

Figure 4.10 shows the differences between two samples that are sent through the spatiotemporal granulation engine. Figure (a) is the sound of fireworks, captured outdoors, while figure (b) is the sound of a choir in a reverberant hall. Notice how there is only a minor difference between the adjacent grains in the choir sample.

Through analysis via spatiotemporal granulation, we have deduced that the spatial resolution of the decomposition, and uniqueness of the spatiotemporal grains is dependent on the following:

1. Order of the microphone array (Spherical Harmonics)
2. Characteristics of the captured signal: spectral (timbre) and temporal characteristics (transient versus sustained sounds)
3. Space where the signal is captured in: reverberant hall versus open space, or acoustically treated spaces



(a) Fireworks (b) Choir

Figure 4.10: Spatiotemporal granulation performed on sources in different spaces. X-Axis= Azimuth (0°- 360°), Y-Axis= Frequency bin, Intensity= Magnitude of bin, Window size = 512 samples

4.3 Transformation

In this section, we discuss the possible transformations that could be applied to the spatiotemporal grains. The transformations that could be applied are not limited to the ones described here. Rather, these are merely starting points, and potentially every transformation that could be applied to classical granulation, could also be applied to spatiotemporal granulation. Furthermore, the extraction of grains in space and time introduces unique novel effects.

4.3.1 Per-grain Transformations

Transformations performed on a "per grain basis", such as per grain reverberation, and per grain filtering, can be applied spatially. For instance, grains that are extracted from a certain direction can be convolved with an Impulse Response that differs from grains that are extracted from another direction.

Another example is to process the grains using a unique bandpass filter, varying in center frequency and Q-value, based on each grain's amplitude. For instance, the grain with the highest energy at a specific time frame could be passed through a filter with a high Q-value, while the adjacent grains could be passed through a filter with a lower Q-value. Alternatively, the center frequency of each adjacent grain could be shifted to create a "spatial filtering" effect.

4.3.2 Granular Substitution

As discussed in Section 4.2.4, the spatial resolution of this technique is greatly dependent on a few factors, including the order of the microphone array used to capture the sound field (in the case of recorded samples). Transient sounds such as fireworks tend to produce unique spatiotemporal grains, compared to long, sustained sounds. Similarly, the sounds captured in an acoustically dry space tends to produce grains that are more distinct, compared to a reverberant space. This translates to the spectral difference of a grain in relation to the adjacent (neighboring) grains. On a first order microphone array, the resulting spatiotemporal grains may be highly correlated.

Granular substitution is the process of substituting grains on each frame, in order to create a more varied texture. The grains on one frame can be substituted with other grains from a different spatial or temporal position, i.e., temporal or spatial index. Additionally, the grains can be substituted with other grains from a completely different spatial (or non spatial) sound recording. The grains for granular substitution can be selected via different techniques, similar to those described in Section 4.4.1.

4.3.3 Dictionary-Based Methods

As an extension to granular decomposition via Dictionary Based Methods [70], Spatiotemporal Granulation can be incorporated to generate sparse approximations of atoms that are localized in time, frequency, and space, resulting in a dictionary which tiles the time-frequency-space plane. Transformations such as morphological filtering (i.e. filtering tailored to specific sound structures), jitter, and mutation (granular morphing/ sonic metamorphosis) could give rise to a new family of transformations that affects temporal, frequency, and spatial domains.

4.3.4 Affine Transformations

Affine transformations such as translate, scale, rotate, shear, and reflect could be performed on the Spatiotemporal slice (Figure 4.11), or the Spatial Read Pointer (Figure 4.12), as discussed in Section 4.4.1.

4.4 Synthesis

As discussed in Section 2.3, the output of a granulation process is greatly dependent on the input signal, and parameters of the granulation engine. In theory, potentially every parameter of classical granulation effects spatiotemporal granulation. As a review, the parameters of classical granulation are as follows:

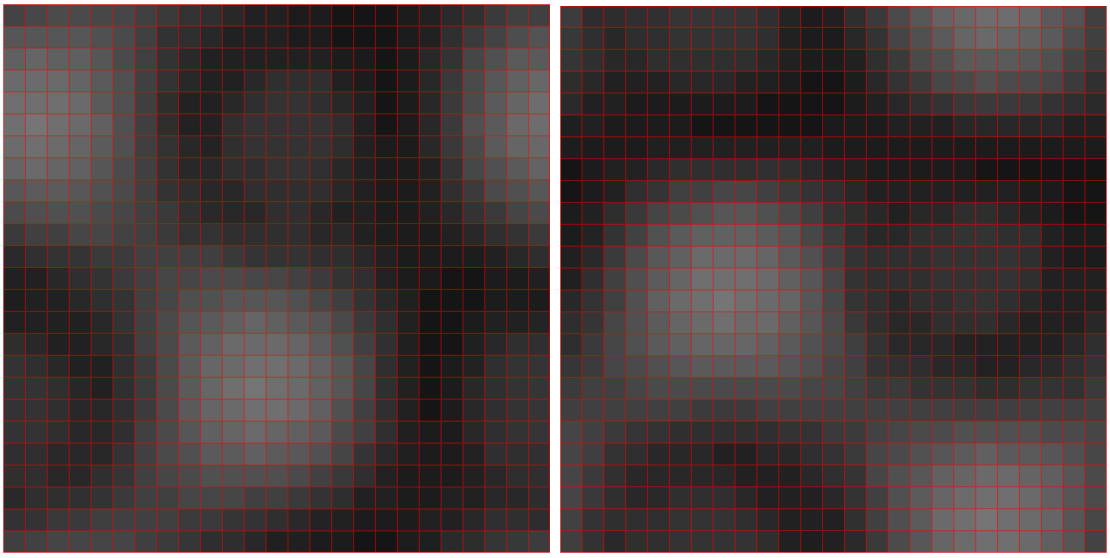


Figure 4.11: Left: Original Spatiotemporal slice, Right: Transformed Spatiotemporal slice– rotate 90° counter clockwise, reflect on Y-axis.

1. Selection order– from input stream: sequential (left to right), quasi-sequential, random (unordered)

2. Pitch transposition of the grains

3. Amplitude of the grains

4. Spatial position of the grains

5. Spatial trajectory of the grains (effective only on large grains)

6. Grain duration

7. Grain density– number of grains per second

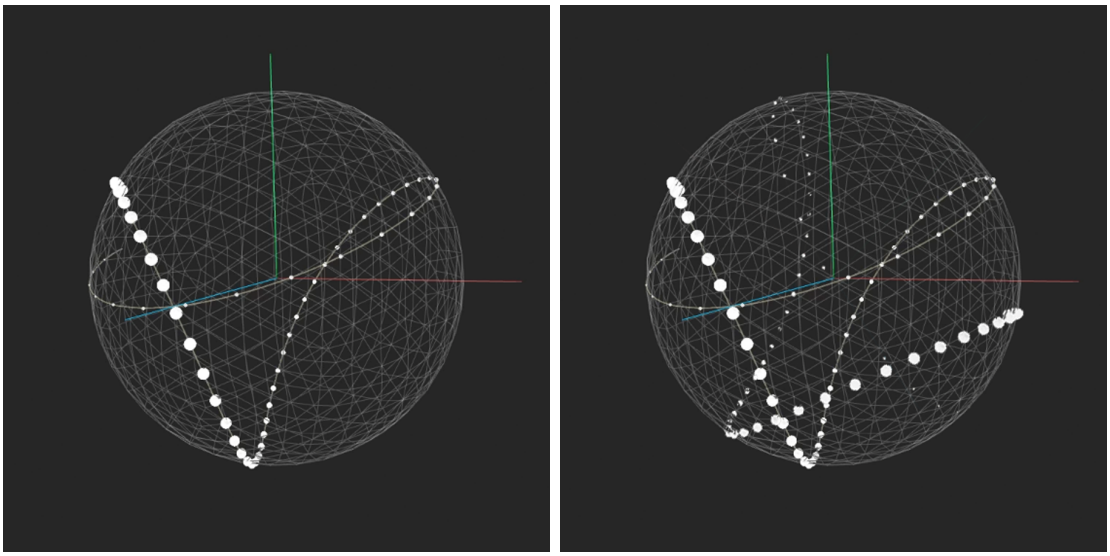


Figure 4.12: Left: Original Spatial Read Pointer, Right: Transformed Spatial Read Pointer– duplicated, and reflected

8. Grain envelope shape

9. Temporal pattern–synchronous or asynchronous

10. Signal processing effects applied on a grain-by-grain basis–filters, reverberators, etc.

4.4.1 Recontextualized Parameter

The following are parameters that acquire a different context through Spatiotemporal Granulation:

Selection Order

The selection order is expanded to include not only the one dimensional selection order in time, but also the spatially encoded layout of the captured sound field. Previous methods of selection are also applicable in the spatial dimension.

Selection Order: Selective Granulation

Specifying only a certain region to be granulated allows us to selectively granulate the sound field. Examples of manipulations that can be performed through selective granulation includes those that are described in Section 4.4.2, Section 4.4.3, and Section 4.4.4. Selection of the areas can be realized via different techniques, including (but not limited to):

1. Regions of the frame, such as quadrants, or sections
2. User defined selection of grains [63]
3. Statistical algorithms for quasi-random selection
4. Audio features

Selection Order: Spatial Read Pointer

In addition to the ability to select single grains (or groups of grains), we implemented a technique to select a sequence of grains, called the "Spatial Read Pointer" (Figure 4.12(a)). Analogous to the read pointer in classical granulation,

the spatial read pointer orbits around a specified trajectory, and extracts grains that fall within the path.

Sudden movement transpositions applied to the continuous trajectory of the spatial read pointer results in quasi-sequential selection. Random triggering of the spatiotemporal grains is achieved by providing the algorithm with a random index in space and time.

To ensure that the spatial read pointer is able to extract grains in the correct spatial position, the orbit needs to be updated at a rate that is at least as high as the trigger rate. This is achieved by calculating the orbit trajectory in the audio callback, at audio rate. As such, not only are the grains extracted from the correct position in space and time, but the movement of the orbit can be increased to audio rate.

The decoupling of temporal and spatial processes allow us to assign two separate unique indexes for the temporal read pointer and the spatial read pointer. This enables us to independently control the movement in space and time, allowing one to potentially simulate relativistic effects. The extreme case would be to freeze time, and "scan" the captured space, which creates an effect that can be thought as *spatially exploring a moment frozen in time*.

Selection Order: Algorithmic/ Generative

As we have shown, the spatial read pointer is one technique for specifying a selection pattern. This functions as a starting point for further investigations in extracting, and triggering spatiotemporal grains. Other algorithms that would be implemented in the near future include fractal based, physics based, statistical, stochastic, and cellular automaton.

Selection Order: Space-time Representation

The representations shown in Figure 4.5 can be thought of as individual slices, or frames of a specific moment in time. If all the frames are successively lined up, we would gain a representation of the full space-time decomposition. This would introduce the possibility of traversing through both dimensions, allowing us to simultaneously extract and synthesize grains from different points in space and time. For example, the Spatial Read Pointer can be used not only to extract grains based on each frame (frozen time), but could also be used to navigate seamlessly to extract any spatiotemporal grain in both dimensions. One could also create a dynamic manipulable cloud containing only spatiotemporal grains extracted from a section of space, in a specific time period.

Selection Order: Spatial Windowing

A temporal grain is created by multiplying a one dimensional audio signal (less than 100 ms) with a specified window (hanning, hamming, blackman, etc). In other words, the algorithm takes a snapshot of a number of samples in time, and removes certain parts of the signal.

Analogous to this process, through spatiotemporal granulation, we can apply a two dimensional window to select only a portion of the signal to be synthesized, i.e, grains to be triggered. In the case of the full space-time representation, a three dimensional window would be used. Properties of the window such as window type can be customized, akin to the temporal domain counterpart [64].

Spatial Position

Spatial position of grains are now dependent on the encoded spatial information. However, we now have the option of decoupling them so as to reassemble the spatiotemporal grains in a different spatial configuration. Previous methods that are described in Section 2.4, as well as other methods of spatializing grains are still applicable.

On the other hand, spatiotemporal granulation introduces a multitude of endless possibilities for spatial transformations. As the encoded sound field is now decomposed into a malleable representation, each grain can be positioned based

on any given rule. For example, one could perform feature analysis of each grain, and spatially (and temporally) group them based on their features, similar to concatenative synthesis [66]. A selection of techniques for spatial transformations are described in Section 4.5.

Spatial Trajectory

The location and motion of sound objects in a spatial scene could now be extracted from a sound field recording. For example, we could extract the motion of a moving object, and impose a different path, or change the course of its original trajectory. Alternatively, we could retain the original motion, but change the content of the sound object. An example of this would be to analyze the motion of a flying insect (such as a bee), and use the extracted path as a trajectory for the sound of a moving train. Additionally, we can spatially disintegrate a sound object based on real-world models, such as smoke or fluid simulation.

As we now have independent control over both the temporal, and the spatial segmentation, we are able to customize and manipulate one domain, without affecting the other. For example, imagine a scene with an exploding sound object, followed by grains traveling outwards in 360° from the location of the initial burst. We can reverse time, causing the sound to be played backwards, and the grains to spatially coalesce, instead of the radial outwards dispersion. Additionally, we

can allow only the spatial dimension to travel backwards, but allow the temporal dimension to progress normally (or vice versa). The perceived effect would resemble the space collapsing into a single point at the origin, but the sound to move through time at normal speed.

Extracting the trajectory of motion allows us to transform and customize the motion of a sound object through processes such as evaporation (sonic disintegration), coalescence [70] (sonic formation), or affine transformations (Section 4.3.4). In addition, it also allows us to map the extracted information to other granulation parameters. For example, we can map the speed of movement to the pitch of the grain, or to the grain's temporal length (or spatial size)– the faster the motion, the smaller the size, or vice versa.

Grain Density

In classical granulation, when we increase the grain density, we allow more grains to overlap, based on the fill factor (Section 2.2.5). The contents of the grain (waveform within the grain) can either be the exact copy, i.e., same temporal selection, or at a different time point in the buffer (including optional transformations).

In the realm of Spatiotemporal Granulation, as each grain contains a different copy of a similar signal (spatial difference), we gain a different effect. The decor-

relation [73, 19] of each individual grain allows us to control the source width of a sound object by manipulating spatial grain density. The control of source width is described in Section 4.5.2.

4.4.2 Spatiotemporal Cross-Synthesis

Cross-synthesis is the process of impressing the characteristics of one signal on to another [37]. Section 4.2.4 discussed about the issues of spatial resolution, and the uniqueness of spatiotemporal grains, while Section 4.3.2 examined the potential of granular substitution– the process of substituting selected grains from one decomposed signal to another.

Spatiotemporal Cross-Synthesis is the process of impressing the characteristics of one sound field recording to another, on a per-grain basis. For example, we can impose the spectral envelope of a spatiotemporal grain on to another grain, causing the original carrier grain to morph into the other. This process allows one to "compose" the space-time slices in order to create a desired space.

Furthermore, this technique allows us to change the spatial characteristics of an encoded signal. As a proof of concept, we have imposed the spatial characteristics of one spatial recording to another spatial recording through the use of each grain's amplitude envelope. The steps that were taken can be summarized as follows:

1. Decompose a spatial sound (Carrier) into grains that are localized in space and time

2. Decompose a different spatial sound (Modulator) into grains that are localized in space and time

3. Calculate the RMS amplitude of Modulator grains

4. Impose the spatial configuration of Modulator grains to Carrier grains

    - Multiply the Interpolated RMS amplitude of Modulator grains with the amplitude of the Carrier grains

5. Encode into Ambisonics B-format

Video example 4.21. Carrier signal 1 (Insects).

Video example 4.22. Modulator signal (Fireworks).

Video example 4.23. Result of Spatiotemporal Cross-synthesis 1.

Video example 4.31. Carrier signal 2 (Organ).

Video example 4.32. Modulator signal (Fireworks).

Video example 4.33. Result of Spatiotemporal Cross-synthesis 2.

Future work involves implementing different types of cross-synthesis algorithms, including convolution and Dictionary Based Methods [70].

4.4.3 Spatiotemporal Stretch

As we are are now dealing with individual grains that are localized in space and time, we can now apply per-grain effects, which results in changing the sound field in its entirety. For example, one could granulate only a single quadrant (of space), and temporally stretch the grains that fall within that area, while allowing the other quadrants to progress at a different time speed. In effect, a moving sound object would retain its spatial trajectory, but assume a different temporal structure as it moves through these selected areas of space.

When a different stretch factor is applied on a per-grain basis, we are in fact warping the space-time continuum, allowing different points in space to traverse at a different temporal speed. In video examples 4.42 and 4.52, we apply a stretch factor that is lower than the original speed (1 / grain spatial index) for each grain. In contrast, a stretch factor that is higher than the original speed (total grains/ grains spatial index) is applied to grains in video examples 4.43 and 4.53.

The perceived pattern appears because the grains are stretched at different factors based on their *spatial index*. The steps that were taken can be summarized as follows:

1. Decompose a spatial sound into grains that are localized in space and time
2. Select spatiotemporal grains to be time stretched (e.g., based on spatial index)
3. Calculate unique stretch factors per spatiotemporal grain
4. Apply granular time stretching for each spatiotemporal grain
    - Contraction: (Stretch factor = 1 / grain spatial index)
    - Expansion: (Stretch factor = total grains / grain spatial index)
5. Encode into Ambisonics B-format

Video example 4.41. Signal 1 (Fireworks).

Video example 4.42. Result of Spatiotemporal Stretch 1: Contraction

Video example 4.43. Result of Spatiotemporal Stretch 1: Expansion

Video example 4.51. Signal 2 (Gamelan).

Video example 4.52. Result of Spatiotemporal Stretch 2: Contraction

Video example 4.53. Result of Spatiotemporal Stretch 2: Expansion

Other ways of spatially selecting these grains would result in a different spatial transformation. For example, if we were to have a moving sound source in the spatial scene, such as a flying bee, we can perform feature analysis and track the spatial movement of the bee (instead of using the spatial index). We can then apply a higher stretch factor for the grains that are closest to the bee, and a lower stretch factor for the other grains.

4.4.4 Spatiotemporal Gate

Gating is the process of allowing only a sound with specific characteristics above a threshold to be played, while removing sounds that are below the threshold. A typical use of this technique can be seen in a *Noise Gate*, where the gate is closed whenever a signal is below a specified threshold. This is to ensure that silence is achieved, as opposed to noise, when there is no signal passing through.

Spatiotemporal Gate allows grains to be played only when it is above a certain threshold. As discussed in Section 4.2.2, we can determine the location of a sound source by analyzing the energy content of each spatiotemporal grain. This process attenuates all the grains that are below a specified threshold, and only allow the grains above a certain threshold to be played. Consequently, this results in spatially narrowing the sound source, and can be thought of an auditory spatial microscope, where we can control the width of the *spatial window*.

In video examples 4.62 and 4.72, the gating parameter is based on each spatiotemporal grains' RMS. The louder grains passes through the gate, while the softer ones become silent. In doing so, we exaggerate the spatial focusness of the sound source, enhancing the localization of the sound source. The steps that were taken can be summarized as follows:

1. Decompose a spatial sound into grains that are localized in space and time
2. Perform feature analysis on the grains as a gating parameter (e.g., RMS)
3. Selectively synthesize grains above a specified threshold
4. Encode into Ambisonics B-format

Video example 4.61. Signal 1 (Fireworks).

Video example 4.62. Result of Spatiotemporal Gate 1.

Video example 4.71. Signal 1 (Gamelan).

Video example 4.72. Result of Spatiotemporal Gate 2.

It is not necessary for the gating parameter to be based on RMS amplitude. Future work involves using other parameters for gating, for example, spatial position, or spectral features (centroid, kurtosis, etc).

4.5 Spatialization

The extracted, and transformed spatiotemporal grains could now be positioned in different locations through the use of various spatialization algorithms and techniques, described in Section 1.6, and Section 3.3.3.

In addition to re-encoding the decomposed sound field using the same order as the original signal, the transformations mentioned in Section 4.3.1 and Section 4.4.4 allow us to perform *upmixing*. As we now have the sound field in the form of quantized spatial positions, and because each spatiotemporal grain now contains a different waveform, we now have the option to re-encode the spatiotemporal grains into a higher order Ambisonic format. Furthermore, we have also explored the possibilities of spatializing the grains using VBAP (Section 3.3.3), and directly assigning the grains to the nearest loudspeaker in the AlloSphere [41].

4.5.1 Exploring Space

In classical granulation, when we freeze the time and extract grains from a static temporal index, the waveform within the grain remains constant. Continuous triggering of the grain results in an output signal that is a repetition of the same grain.

In the realm of spatiotemporal granulation, we now have the ability to explore the spatial dimension of sounds, independent from the temporal dimension. The

temporal index and the spatial index are independent of each other, and can progress using its own unique pattern. An instance of this is shown in Section 4.4.1.

The extreme case would be to freeze time, and "scan" the captured space, which would result in *spatially exploring a moment frozen in time.* Once the time domain is frozen, the spatial read pointer and various other algorithms such as those described in Section 4.4.1 could be used to extract and trigger the grains. In contrast to the outcome of a similar process applied to classical granulation, we now have different waveforms for each grain based on the spatial index– that is to say, each grain contains a waveform corresponding to the sound from different locations in the original encoded sound field.

4.5.2 Spatial Stretch

Analogous to granular time stretching processes, spatial stretching allows us to control the source width, and smear the spatial position of a sound object. This is achieved by increasing the grain density (Section 2.2.5), and overlapping the decorrelated grains in a specific position. The decorrelation of grains can be achieved via a variety of transformations, such as per-grain transformations listed in Section 4.3.1, or random modulation of sinusoidal components [19]. The process of decorrelating grains, and resynthesizing the sound field could potentially reduce

comb filtering effects when a lower order Ambisonics signal is decoded over a large number of loudspeakers.

Spatial stretching, in addition to temporal stretching can be used to transform a noisy spatial scene into an ambient-like environment. For example, the discrete grains from a noisy recording can be transformed into an enveloping ambient space by spatially stretching, and overlapping the spatiotemporal grains.

4.5.3 Spatial Warp

The spatiotemporal grains in a given frame (Figure 4.5) can be rearranged, and concentrated in a particular area (spatial density), or spread across a particular region of the frame. By controlling the spatial density over time, we are able to simulate the effect of warping space, without affecting the temporal domain of the sound material. Methods of selection for the grains to be controlled are similar to those described in Section 4.4.1.

4.5.4 Spatial Descriptor

The analysis of discretized grains in space and time could lead to the possibility of spatial audio descriptors. For example, one could analyze the spatiotemporal grains of a single frame, and determine the spatial centroid of a given space. The spatial features could also be combined with other temporal, or spectral features,

such as spectral spread, skewness, kurtosis, harmonic and noise energy, which would allow us to measure the spatial distribution of specific features.

By analyzing the spatial scene, we are able to spatially segregate sound sources based on their location. This could lead to the potential of instrument separation/ extraction via spatial properties– Spatial Source Separation. For example, we could analyze the position of specific instruments/ performers in a spatial recording, and separate the instruments based on their spatial location, in addition to spectral qualities.

Furthermore, this information could also be used as a template to map non-spatial recordings (of performances). An example case would be to train a machine learning algorithm with a database of instruments placed around a sound field microphone for performances. We can then feed the system with instrument tracks, and request the algorithm to spatialize the input sounds based on the trained data.

Chapter 5

Implementation: Angkasa

The word Angkasa originates from the Malay language, derived from the Sanskrit term *Ākāśa*. Although the root word bears various levels of meaning, one of the most common translation refers to *space*.

In the context of our research, Angkasa is a software tool that allows a user to perform Spatiotemporal Granulation. This includes the analysis, transformation, synthesis, and spatialization processes described in Chapter 4. The software is designed to be used as a creative tool for composition, real-time musical instrument, or as an analytical tool.

5.1 Prototype

Initial experiments to determine the validity of Spatiotemporal Granulation were realized using Python [5], in particular the interactive IPython notebook [48]. Early explorations include decomposition of ambisonic files (Section 4.2.1) into spatiotemporal grains (Figures 4.4, 4.5), and spectral analysis (Section 4.2.2) of the grains.

As the analysis proved that the segmentation of space creates spectrally unique spatiotemporal grains, it soon became necessary to acoustically verify the theory. We built an interactive system using Max/ MSP [3] (Figure 5.1), and the result proved that the spatiotemporal grains does sound different from one another, even for a first order ambisonics signal.

However, Max/ MSP soon proved to be a limited solution, due to the inherent limitation that control data is not processed as often as signal data, i.e., if a control data is triggered at a rate that is higher than the processed rate, then it is not processed within the correct audio block. Since granular synthesis involves triggering grains that are less than 100 ms, many control data will not be triggered correctly in time.

There are various ways to overcome this issue, such as reducing the defined block size, and using sample-accurate trigger externals. However, we chose to move away from this environment, not only due to the described limitation, but

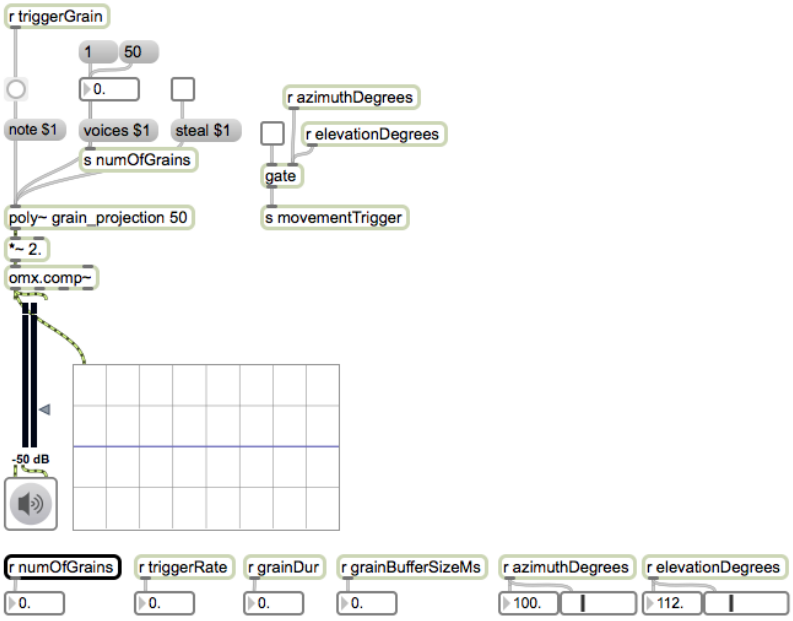


Figure 5.1: Prototype in Max/MSP

also to find a more suitable platform to implement specific transformations, and real-time visualization.

5.2 First Iteration

The first iteration of Angkasa was built using openFrameworks [4] (C++ toolkit) on a 2015 Mac Pro (OSX 10.10.5). It features the extraction of spatiotemporal grains in space and time, which could be performed using a variety of techniques listed below. The software also allowed a user to reorganize the

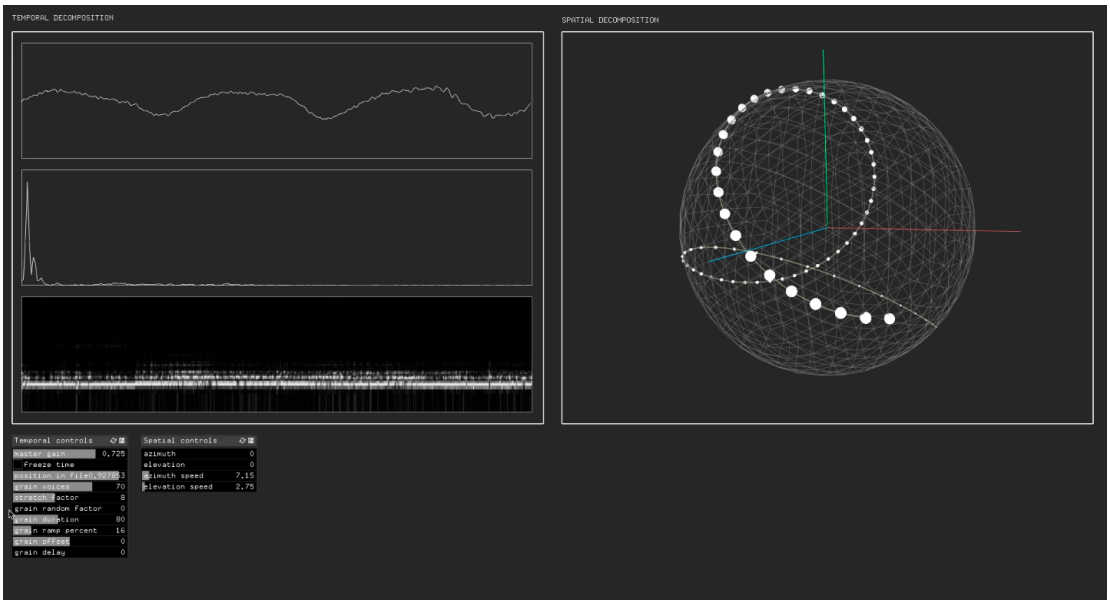


Figure 5.2: GUI for first iteration of Angkasa

grains into a new temporal re-assembly, as well as a simple one channel spatial re-assembly.

The development phase was primarily carried out during the *Space-Media-Sound*(1.8) exchange program in Karlsruhe. The software is open source, and can be downloaded from `https://github.com/muhammadhafiz/angkasa`.

Video example 5.1. Screen capture of Angkasa.

5.2.1 Interface

The following temporal control parameters are accessible via the graphical user interface:

- Position in file
- Freeze time (static temporal window)
- Grain voices
- Stretch factor
- Random factor
- Duration
- Window type
- Offset
- Delay

Spatial location where the spatiotemporal grains are extracted from is determined by specifying a value for azimuth & elevation. The spatial index can be set using the following techniques:

1. Independent GUI sliders for azimuth and elevation
2. Point picker
3. Algorithmically (discussed in Chapter 4.4.1)

5.2.2 Visualization

Visualization of the temporal decomposition includes temporal, and frequency domain plots, as well as a spectrogram to monitor the extracted grains in real time (Figure 5.2– top left). The temporal and spatial parameters listed above are controllable via GUI sliders.

The spatial domain is visually depicted as a geodesic sphere, which represents the captured sound field. The spatiotemporal grains are visualized as smaller spheres on the surface of the geodesic sphere (Figure 5.2– top right), corresponding to the location where the grains are extracted from.

5.3 Second Iteration

The first iteration of Angkasa described above allowed a user to analyze, and synthesize spatiotemporal grains. However, it did not address the process of transformations, and pluriphonic spatial re-construction. The second iteration of Angkasa was developed to address these processes. In order to contrast the original recorded sound field and the transformed result, the software needs to be spatialized in a multichannel setup.

The AlloSphere (Figure 5.3) is a spherical space in which immersive, virtual environments allow users to explore large-scale data sets through multimodal, in-

teractive media [7, 41, 21]. Housed in the California NanoSystems Institute, the three-story high structure allow sounds to be spatialized via 54 Meyer Sound MM-4XP compact loudspeakers, and a 2250-watt Meyer 600-HP subwoofer. Figure 5.4 shows the speaker configuration for the AlloSphere. Additionally, the AlloSphere also provides 360° real-time stereographic visualization using a cluster of servers driving 26 high-resolution projectors.

Video example 5.2. Video documentation of Angkasa in the AlloSphere.

This version of Angkasa was developed in the beginning of the Fall 2016 quarter, and was completed at the end of the Winter 2017 quarter. It was written in AlloSystem [1], a cross-platform suite of C++ components for building interactive multimedia tools and applications. The software is open source, and can be downloaded from `https://github.com/muhammadhafiz/angkasa.allosphere`. On March 17th 2017, the author performed Angkasa in the AlloSphere (Figure 5.5) as part of the Doctoral defense. On May 19th 2017, the author gave five performances as part of the MAT End of the Year Show (`http://show.mat.ucsb.edu/`).

The second iteration of Angkasa includes a few additional features, namely:

- Extraction of spatiotemporal grains in space and time

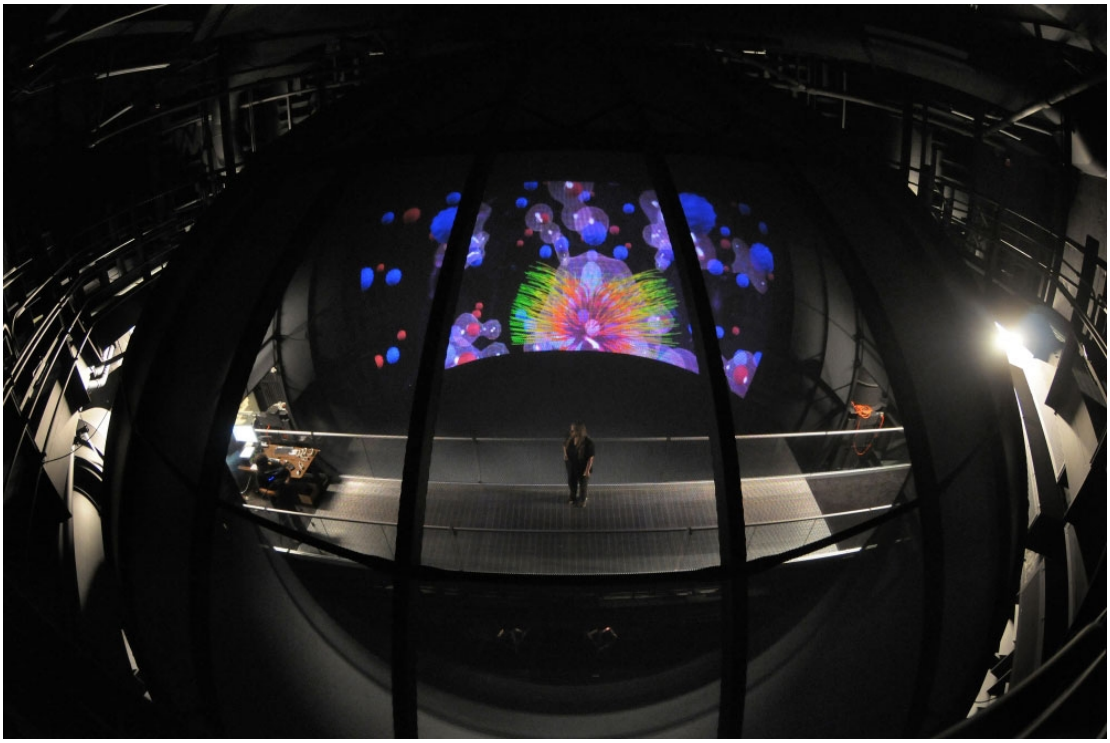


Figure 5.3: External view of the AlloSphere. Image from the California NanoSystems Institute

  - Patterns for extraction: Spatial read pointer
  - Patterns for extraction: Point picker
  - Patterns for extraction: Stochastic algorithm

- Spatial transformation
  - Spatiotemporal Cross-synthesis (Section 4.4.2)
  - Spatiotemporal Stretch (Section 4.4.3)
  - Spatiotemporal Gate (Section 4.4.4)

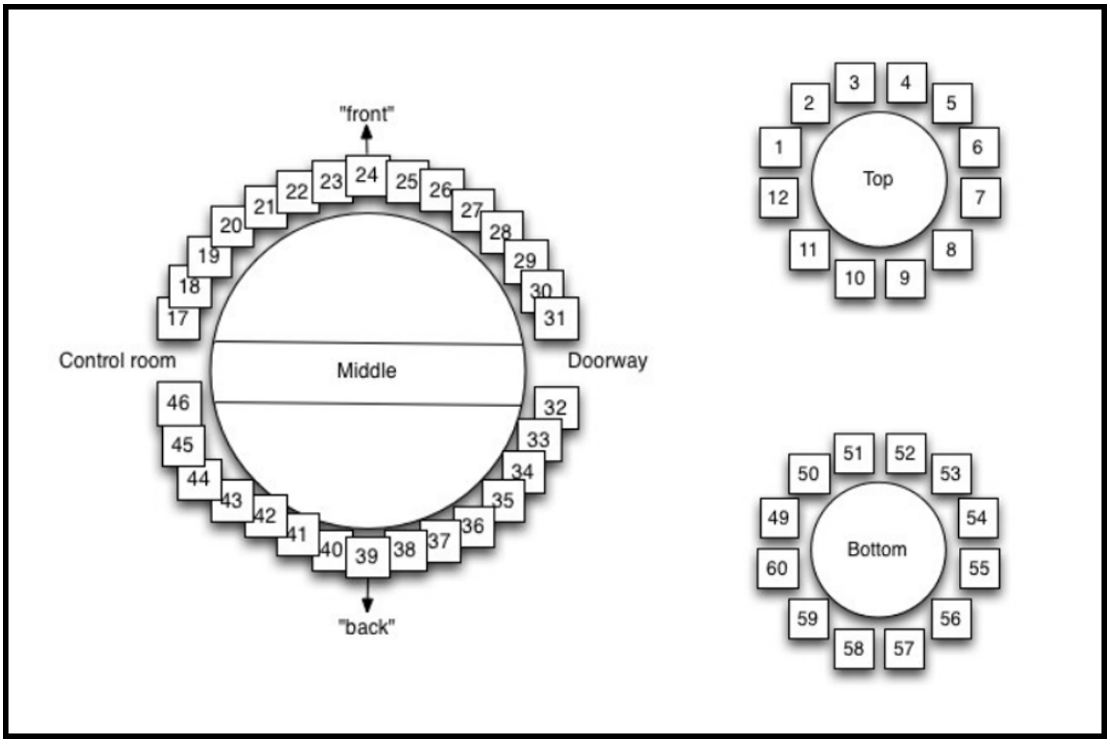


Figure 5.4: AlloSphere loudspeaker configuration

- Temporal re-assembly
- Multi channel spatial re-assembly in the AlloSphere
  - Re-encoding into original format
  - Encoding into higher order Ambisonics
  - Spatialization via Ambisonics
  - Spatialization via VBAP
  - Assignment of Spatiotemporal Grains to closest speaker

Figure 5.5: Angkasa in the AlloSphere on March 17th 2017

5.3.1 Interface

Figure 5.6 is a screenshot of the Graphical User Interface for Angkasa. In the performances mentioned above, the control interface was placed in the center of the AlloSphere, which can be seen in Figure 5.5. In the center of the GUI, every decomposed spatiotemporal grain is positioned corresponding to its extracted position in space. On the left side, the interface displays sliders with the current settings for each parameter, while the right portion of the interface shows the component signals for the re-encoded Ambisonics output.

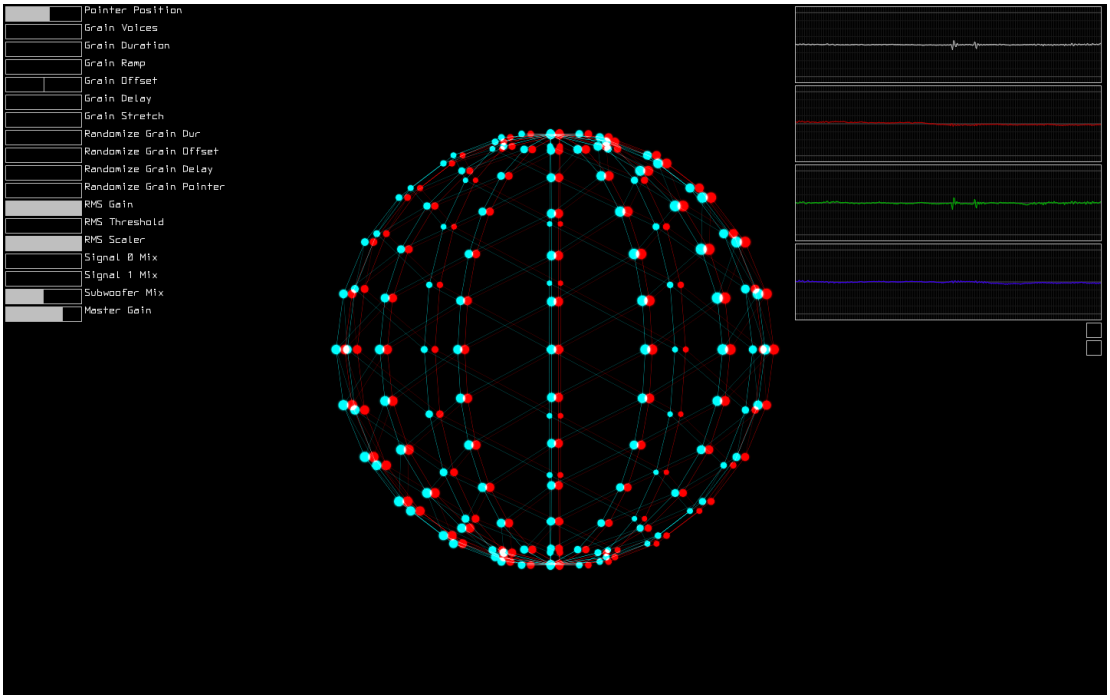


Figure 5.6: Interface for Angkasa in the AlloSphere

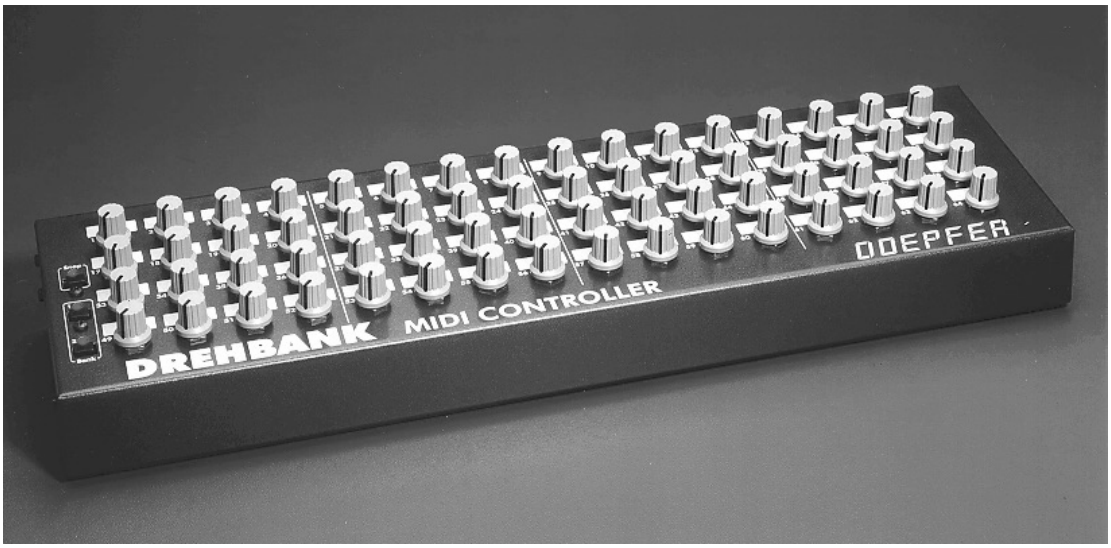


Figure 5.7: Doepfer Drehbank [2]

The parameters can be controlled via MIDI, external OSC [78] controllers, or other types of physical controllers. Currently, the parameters are controlled using the Doepfer Drehbank (`http://www.doepfer.de/db.htm`), a programmable 64 knob MIDI controller (Figure 5.7). The available control parameters are as follows:

- Grain Voices
- Grain Duration
- Grain Ramp
- Grain Offset
- Grain Delay
- Grain Stretch
- RMS Gain
- RMS Threshold
- RMS Scaler
- Subwoofer Mix
- Master Gain
- Randomize Grain Duration
- Randomize Grain Offset
- Randomize Grain Delay
- Randomize Grain Pointer
- Signal Mix 0

- Signal Mix 1

5.3.2 Visualization

As opposed to the flat visualization in the first iteration of Angkasa, the second iteration's visualization plays a more prominent role. Each spatiotemporal grain is visually positioned on the surface of the AlloSphere, accompanying its sonic counterpart. The size and intensity of the visual grains correspond to its amplitude (per grain RMS amplitude). Figure 5.8 shows a flat rendering, i.e., overall exterior view of what is rendered on the surface of the AlloSphere, as can be seen in Figure 5.5.

The visuals are rendered using AlloSphere's stereo display, giving the perception of depth, which can be mapped to a different parameter. The closely correlated visual and auditory stimuli assists each spatial event to be perceptually coherent, giving a better sense of localization [64, 18]

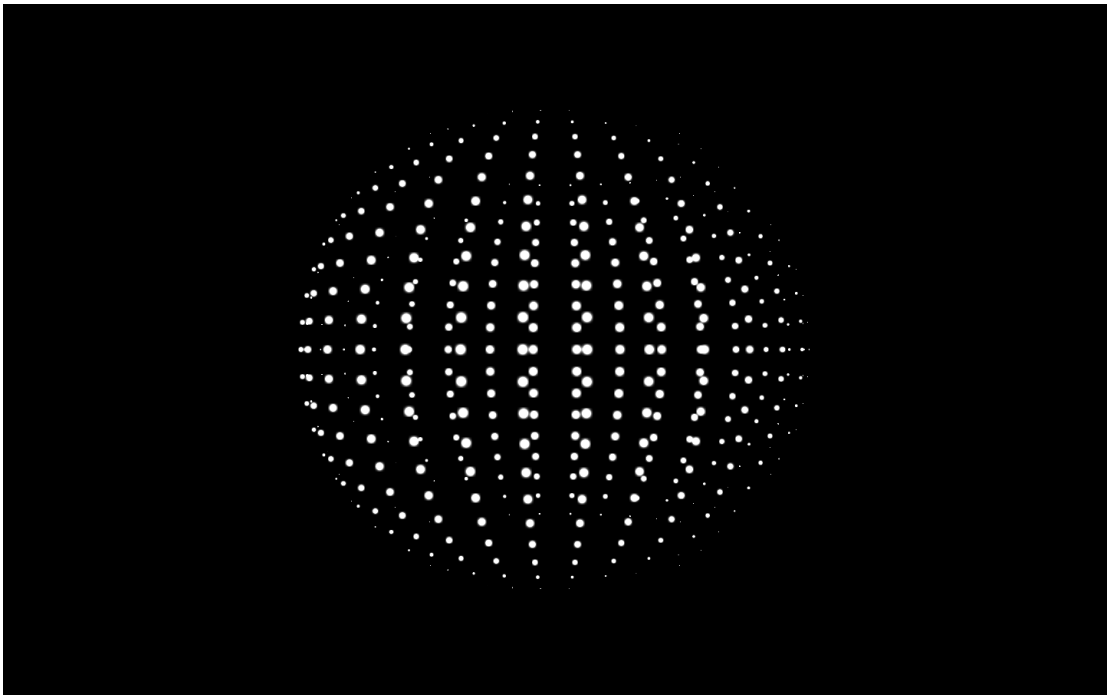


Figure 5.8: Visualization of Spatiotemporal Grains in the AlloSphere

Chapter 6

Conclusion

This dissertation introduces a novel theory and technique called Spatiotemporal Granulation. The underlying foundation of this theory is the spatial and temporal segmentation of spatially encoded signals to produce grains that are localized in both space and time– Spatiotemporal Grains. This malleable representation enables space to become an additional expressive parameter through the process of maintaining, breaking or contrasting the original space with the granulated space.

We presented a brief overview of microsound and spatial sound as a platform to contextualize this research, and outline the issues and current practices within the domains. The validity of this theory is assessed/verified through experiments that are carried out using our software implementations, described in Chapter 5.

The use of Angkasa is not only limited to analytical applications, such as those that are presented in Section 4.2. The software can also been used as a creative tool for computer music performances, such as those that were carried out in the AlloSphere.

As a review, the three main experiments used to verify the theory are as follows:

1. **Analysis:**

    1.1. Decompose spatial sound, and prove that the spatiotemporal grains are unique in time, frequency, and space

2. **Transformation:**

    2.1. Extract the spatiotemporal grains, and selectively trigger them in any arbitrary order

3. **Synthesis:**

    3.1. Reassemble the grains into a new spatial and temporal pattern (manipulate the spatial sound's encoded configuration)

        3.1.1. Resynthesis: Reconstruction from extracted grains

        3.1.2. Complex multi-channel reassembly

6.1 Results

1. **Analysis**: In Section 4.2, we have proven that the encoded sound field can be segmented into grains that are localized in space and time. Each spatiotemporal grain is unique based on a few factors, listed in Section 4.2.4. In Section 4.2.2, we prove that the frequency component is retained within each spatiotemporal grain. Although we have not fully explored the use of the grains' spectral information, suffice to say that various features can be extracted from this representation, which could potentially be used for further transformations.

2. **Transformation**: Section 4.4.1 presents a number of techniques to extract the spatiotemporal grains, and synthesize the grains in a different order. Both iterations of Angkasa demonstrates the techniques that were discussed.

3. **Synthesis**: In Section 4.2.3, we prove that the decomposed grains can be used to reconstruct the original signal. This allows us to perform various transformations (Section 4.3) in between the analysis and synthesis stages. Section 4.4.1 and Section 4.5.1 presents techniques on how the decomposed grains can be reassembled into a new space and time pattern.

   Manipulation of the spatial sound's encoded configuration via Spatiotemporal Cross-Synthesis, Spatiotemporal Stretch, and Spatiotemporal Gate

are discussed in Section 4.4.2, Section 4.4.3, and Section 4.4.4. The AlloSphere version of Angkasa allows a user to perform these techniques, and manipulate the encoded configuration in real-time. Furthermore, complex multi-channel reassembly is made possible through the use of spatialization techniques presented in Section 4.5. Additionally, the reconstructed spatial scene is also rendered into binaural format.

6.2 Future Work

Development of this research will proceed in different areas, including (but not limited to) extraction, analysis, control, transformation, synthesis, spatialization, and visualization of spatiotemporal grains. Examples of research directions include:

- Audiovisual Spatiotemporal Granulation: Using 360° camera
- Micro Manipulation of sound trajectory: Change the trajectory of a moving sound object
- Manipulation of grains based on feature analysis
- Rearrange spatiotemporal grains based on features
- Spatial source separation
- Usage of Higher Order Ambisonic recordings
- Synthesize the sound field, in combination with sound field recordings [9]

Appendix A

The Angkasa Program

Documentation and code for the first iteration of Angkasa (Section 5.2) can be downloaded from:

`https://github.com/muhammadhafiz/angkasa`

Documentation and code for the AlloSphere version of Angkasa (Section 5.3) can be downloaded from:

`https://github.com/muhammadhafiz/angkasa.allosphere`

Appendix B

Audio and Video Examples

The audio and video examples used in this document can be downloaded from:

`https://github.com/muhammadhafiz/spatiotemporal_granulation`

Audio example 3.1. Result of Tape Echo Feedback.

Audio example 4.1. B-Format Fireworks.

Audio example 4.2. B-Format Choir.

Audio example 4.3. B-Format Steamtrain.

Audio example 4.4. B-Format Gulls.

Audio example 4.5. B-Format Buzzard.

Video example 4.1. Result of detecting source direction by estimating the peak of spectral bins.

Video example 4.21. Carrier signal 1 (Insects).

Video example 4.22. Modulator signal (Fireworks).

Video example 4.23. Result of Spatiotemporal Cross-synthesis 1.

Video example 4.31. Carrier signal 2 (Organ).

Video example 4.32. Modulator signal (Fireworks).

Video example 4.33. Result of Spatiotemporal Cross-synthesis 2.

Video example 4.41. Signal 1 (Fireworks).

Video example 4.42. Result of Spatiotemporal Stretch 1: Contraction

Video example 4.43. Result of Spatiotemporal Stretch 1: Expansion

Video example 4.51. Signal 2 (Gamelan).

Video example 4.52. Result of Spatiotemporal Stretch 2: Contraction

Video example 4.53. Result of Spatiotemporal Stretch 2: Expansion

Video example 4.61. Signal 1 (Fireworks).

Video example 4.62. Result of Spatiotemporal Gate 1.

Video example 4.71. Signal 1 (Gamelan).

Video example 4.72. Result of Spatiotemporal Gate 2.

Video example 5.1. Screen capture of Angkasa.

Video example 5.2. Video documentation of Angkasa in the AlloSphere.

Bibliography

[1] AlloSystem. Cross-platform suite of C++ components for building interactive multimedia tools and applications. https://github.com/AlloSphere-Research-Group/AlloSystem.

[2] Doepfer Drehbank. Programmable universal MIDI controller. http://www.doepfer.de/db.htm.

[3] Max/MSP. Visual Programming Language. https://cycling74.com/products/max/.

[4] openFrameworks. Open source toolkit for creative coding. http://openframeworks.cc/.

[5] Python Software Foundation. Python Language Reference, version 2.7. https://www.python.org/.

[6] Visual representation of Ambisonic components. https://en.wikipedia.org/wiki/Ambisonics#/media/File:Spherical_Harmonics_deg3.png.

[7] Xavier Amatriain, JoAnn Kuchera-Morin, Tobias Hollerer, and Stephen Travis Pope. The allosphere: Immersive multimedia for scientific discovery and artistic exploration. *IEEE Multimedia*, 16(2):64–75, 2009.

[8] M. Baalman. Spatial composition techniques and spatialisation technologies. *Organised Sound*, 15(3):209–218, 2010.

[9] Natasha Barrett. The perception, evaluation and creative application of high order ambisonics in contemporary music practice, 2012.

[10] Isaac Beekman. Journal tenu par isaac beekman de 1604 à 1634. Four volumes, 1953.

[11] Leo Beranek. *Concert Halls and Opera Houses— Music, Acoustics, and Architecture*. Springer-Verlag New York, 2 edition, 2004.

[12] Leo Beranek. Riding the waves—a life in sound, science, and industry. *The Journal of the Acoustical Society of America*, 123(4):1817–1818, 2008.

[13] A. J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36:977–995, 1988.

[14] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2779, 1993.

[15] Tim Blackwell and Michael Young. *Swarm Granulator*, pages 399–408. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[16] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.

[17] Edwin G. Boring. Auditory theory with special reference to intensity, volume, and localization. *The American Journal of Psychology*, 37(2):157–188, 1926.

[18] Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 1994.

[19] Andres Cabrera. *Control of Source Width in Multichannel Reproduction Through Sinusoidal Modeling*. PhD thesis, Queen's University Belfast, 2012.

[20] Andrés Cabrera and Gary Kendall. Multichannel control of spatial extent through sinusoidal partial modulation. In *Proceedings of the Sound and Music Computing Conference*, pages 532–537, Stockholm, Sweden, 2013. Logos Verlag Berlin.

[21] Andrés Cabrera, JoAnn Kuchera-Morin, and Curtis Roads. The evolution of spatial audio in the allosphere. *Computer Music Journal*, 40(4):47–61, 2016.

[22] Christian Clozier. Composition-diffusion/interprétation en musique électroacoustique. In F. Barrière and G. Bennett, editors, *Composition/Diffusion en Musique Electroacoustique*, pages 52–101, Bourges, 1998. Éditions Mnémosyne.

[23] Christian Clozier. The gmebaphone concept and the cybérnaphone instrument. *Computer Music Journal*, 25(4):81–90, 2001.

[24] William Cochran. *The Dynamics of Atoms in Crystals*. London: Edward Arnold, 1973.

[25] S. Conti, P.Roux, D.Demer, and J.Rosny. Let's hear how big you are. `http://acoustics.org/pressroom/httpdocs/146th/Conti.html`, 2003.

[26] H. G. Fisher and S. J. Freedman. Localization of sound during simulated unilateral conductive hearing loss. *Acta Oto-Laryngologica*, 66(1-6):213–220, 1968.

[27] Dennis Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, 93(3):429–457, 1946.

[28] Dennis Gabor. Acoustical quanta and the theory of hearing. *Nature*, 159(1044):591–594, 1947.

[29] Dennis Gabor. Lectures on communication theory. *Technical Report 238, Research Laboratory of Electronics*, 1952.

[30] Mrinalkanti Gangopadhyaya. *Indian Atomism: History and Sources*. Atlantic Highlands, New Jersey: Humanities, Bagchi indological series, 1981.

[31] Martin Gardner. *Fads and Fallacies in the Name of Science*. New York: Dover, 1957.

[32] William G. Gardner. Efficient convolution without input-output delay. *J. Audio Eng. Soc*, 43(3):127–136, 1995.

[33] Michael A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc*, 21(1):2–10, 1973.

[34] Michael A. Gerzon. General metatheory of auditory localisation. In *Audio Engineering Society Convention 92*, Mar 1992.

[35] David Griesinger. Objective measures of spaciousness and envelopment. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, Mar 1999.

[36] Paul M. Hofman, Jos G. A. Van Riswick, and A. John Van Opstal. Relearning sound localization with new ears. *Nat Neurosci*, 1(5):417–421, 09 1998.

[37] Julius O. Smith III. Cross-synthesis. `https://ccrma.stanford.edu/~jos/sasp/Cross_Synthesis.html#26450`.

[38] Gary S. Kendall. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4):71–87, 1995.

[39] Kohichi Kurozumi and Kengo Ohgushi. The relationship between the cross-correlation coefficient of twochannel acoustic signals and sound image quality. *The Journal of the Acoustical Society of America*, 74(6):1726–1733, 1983.

[40] Trond Lossius, Pascal Baltazar, and Théo de la Hogue. DBAP - distance-based amplitude panning. In *Proceedings of the 2009 International Computer Music Conference, ICMC 2009, Montreal, Quebec, Canada, August 16-21, 2009*, 2009.

[41] JoAnn Kuchera-Morin; Matthew Wright; Graham Wakefield; Charles Roberts; Dennis Adderton; Behzad Sajadi; Tobias Höllerer; Aditi Majumder. Immersive full-surround multi-user system design. *Computers & Graphics*, 40:10–21, May 2014.

[42] Dave G. Malham. *Space in Music - Music in Space*. PhD thesis, University of York, 2003.

[43] James McCartney. Supercollider, a new real time synthesis language. In *Proceedings of International Computer Music Conference*, Hong Kong, 1996.

[44] Aaron McLeran, Curtis Roads, Bob L. Sturm, and John J. Shynk. Granular sound spatialization using dictionary-based methods. In *Proceedings of the 5th Sound and Music Computing Conference*, Berlin, Germany, 2008.

[45] A. Mills. On the minimum audible angle. *Acoustical Society of America Journal*, 30:237, 1958.

[46] M. Morimoto and Z. Maekawa. Effects of low frequency components on auditory spaciousness. *Acustica*, 66:190–196, 1988.

[47] Markus Noisternig, Alois Sontacchi, Thomas Musil, and Robert Holdrich. A 3D ambisonic based binaural sound reproduction system. *AES 24th International Conference on Multichannel Audio*, 2003.

[48] Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[49] D. R. Perrott and T. N. Buell. Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise. *Acoustical Society of America Journal*, 72:1413–1417, November 1982.

[50] David R. Perrott, Alan Musicant, and Bettina Schwethelm. The expanding-image effect: The concept of tonal volume revisited. *Journal of Auditory Research*, 20(1):43–55, 1980.

[51] Guillaume Potard and Ian Burnett. A study on sound source apparent shape and wideness. In *Proceedings of International Conference on Auditory Display*, pages 25–28, Boston, USA, 2003. Boston University Publications.

[52] Miller Puckette. *The Theory and Technique of Electronic Music*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007.

[53] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc*, 45(6):456–466, 1997.

[54] Lawrence R. Rabiner and Bernard Gold. Theory and application of digital signal processing. *Physik in unserer Zeit*, 7(4):126–127, 1976.

[55] Lord Rayleigh. On our perception of the direction of a source of sound. *Proceedings of the Musical Association*, 2:pp. 75–84, 1875.

[56] Lord Rayleigh. *The Theory of Sound*. Dover Publications, 1945.

[57] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *SIGGRAPH Computer Graphics*, 21(4):25–34, August 1987.

[58] Curtis Roads. Asynchronous granular synthesis. In Giovanni De Poli, Aldo Piccialli, and Curtis Roads, editors, *Representations of Musical Signals*, pages 143–186. MIT Press, Cambridge, MA, USA, 1991.

[59] Curtis Roads. *Microsound*. MIT Press, 2001.

[60] Curtis Roads. *Composing Electronic Music: A New Aesthetic*. Oxford University Press, New York, NY, USA, 2015.

[61] Curtis Roads, Aldo Piccialli, Giovanni De Poli, and Stephen T. Pope, editors. *Musical Signal Processing*. Routledge, London, 1997.

[62] Muhammad Hafiz Wan Rosli and Andres Cabrera. Application of gestalt principles to multimodal data representation. IEEE VIS, 2014.

[63] Muhammad Hafiz Wan Rosli and Andres Cabrera. Gestalt principles in multimodal data representation. *Computer Graphics and Applications, IEEE*, 35(2):80–87, Mar 2015.

[64] Muhammad Hafiz Wan Rosli, Andres Cabrera, Matthew Wright, and Curtis Roads. Granular model of multidimensional spatial sonification. Maynooth, Ireland, 2015. Sound and Music Computing.

[65] Robert Sazdov, Garth Paine, and Kate Stevens. *Perceptual investigation into envelopment, spatial clarity, and engulfment in reproduced multi-channel audio*. Audio Engineering Society, 2007.

[66] Diemo Schwarz. A system for data-driven concatenative sound synthesis. In *Digital Audio Effects (DAFx)*, Verona, Italy, 2000.

[67] Dennis Smalley. Spatial experience in electro-acoustic music. pages 123–126, 1991.

[68] W. Snow. Basic principles of stereophonic sound. *IRE Transactions on Audio*, AU-3(2):42–53, March 1955.

[69] S. S. Stevens and E. B. Newman. The localization of actual sources of sound. *The American Journal of Psychology*, 48(2):297–306, 1936.

[70] Bob L. Sturm, Curtis Roads, Aaron McLeran, and John J. Shynk. Analysis, visualization, and transformation of audio signals using dictionary-based methods. In *Proceedings of International Computer Music Conference*, Belfast, 2008.

[71] F.M. Tonning. Directional audiometry: II. the influence of azimuth on the perception of speech. *Acta Oto-Laryngologica*, 72(1-6):352–357, 1971.

[72] Ryan H. Torchia and Cort Lippe. Techniques for multi-channel real-time spatial distribution using frequency-domain processing. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, NIME '04, pages 116–119, Singapore, Singapore, 2004. National University of Singapore.

[73] Barry Truax. Composition and diffusion: space in sound in space. *Organised Sound*, 3:141–146, Aug. 1998.

[74] Horacio Vaggione. Jeux d'espaces: conjonctions et disjonctions. pages 119–124, 1991.

[75] Horacio Vaggione. Articulating micro-time. *Computer Music Journal*, 20(2):33–38, 1996.

[76] E. Verheijen. *Sound Reproduction Via Wave Field Synthesis*. PhD thesis, Delft University of Technology, Delft, NL, 1997.

[77] Scott Wilson. Spatial swarm granulation. In *Proceedings of International Computer Music Conference*, Belfast, 2008.

[78] Matthew Wright and Adrian Freed. Open sound control: A new protocol for communicating with sound synthesizers. Thessaloniki, Hellas, 1997. International Computer Music Association.